

TESIS DOCTORAL

CARCINOMA DE TIROIDES DE ESTIRPE FOLICULAR:
IDENTIFICACIÓN Y CARACTERIZACIÓN
FUNCIONAL DE GENES DE BAJA PENETRANCIA



IÑIGO LANDA LÓPEZ
MADRID 2011



TESIS DOCTORAL

**Carcinoma de tiroides de estirpe folicular:
identificación y caracterización funcional
de genes de baja penetrancia**

IÑIGO LANDA LÓPEZ

Madrid, 2011

Diseño de portada: Urko Errazquin, a partir de "Paisaje de mariposas" de Salvador Dalí (1957)



Centro Nacional
de Investigaciones Oncológicas

Programa de Genética del Cáncer Humano
Grupo de Cáncer Endocrino Hereditario

Directora de Tesis
Dra. Mercedes Robledo

La Dra. Mercedes Robledo Batanero, jefe del Grupo de Cáncer Endocrino Hereditario del Centro Nacional de Investigaciones Oncológicas,

CERTIFICA que Don Iñigo Landa López ha realizado bajo su dirección la Tesis Doctoral que lleva por título *"Carcinoma de tiroides de estirpe folicular: identificación y caracterización funcional de genes de baja penetrancia"*.

Tras revisar el presente trabajo, considera que reúne la suficiente calidad para su presentación y defensa.

En Madrid, a 3 de marzo de 2011.



VºBº de la Directora de la Tesis

Mercedes Robledo

Estimados Señores,

Como jefa del grupo de Cáncer Endocrino Hereditario del Centro Nacional de Investigaciones Oncológicas y profesora honoraria de la Universidad Autónoma de Madrid, certifico que Don Iñigo Landa López ha realizado bajo mi dirección la Tesis Doctoral que lleva por título: "*Carcinoma de tiroides de estirpe folicular: identificación y caracterización funcional de genes de baja penetrancia*". Los resultados de esta Tesis relativos a la identificación de factores genéticos de riesgo que contribuyen a la susceptibilidad de desarrollar cáncer de tiroides derivados de célula folicular, han sido presentados en varios **Congresos Nacionales e Internacionales**:

Comunicaciones Orales:

1: **Landa I**, Ruiz-Llorente S, Montero-Conde C, Inglada-Pérez L, Schiavi F, Leskelä S, Pita G, Milne R, Maravall J, Ramos I, Andía V, Rodríguez-Poyo P, Jara-Albarrán A, Meoro A, del Peso C, Arribas L, Iglesias P, Caballero J, Serrano J, Picó A, Pomares F, Giménez G, López-Mondéjar P, Castello R, Merante-Boschin I, Pelizzo MR, Mauricio D, Opocher G, Rodríguez-Antona C, González-Neira A, Matías-Guiu X, Santisteban P, Robledo M. **A two-step candidate-gene association study identifies FOXE1 as a low penetrance gene associated to thyroid cancer and reveals the underlying mechanism.** 34th Annual Meeting of the European Thyroid Association. September 5th-9th 2009, Lisbon (Portugal). ****Selected for the "Topic Highlights Session".***

2: **Landa I**, Ruiz-Llorente S, Montero-Conde C, Leandro-García LJ, Leskelä S, López-Jiménez E, Maliszewska A, Inglada-Pérez L, De La Vega L, Pita G, Alonso

MR, Maravall J, Andía V, Álvarez-Escolá C, Meoro A, Caballero J, Blanco C, Díaz-Pérez J, Serrano J, Mauricio D, Cascón A, Rodríguez-Antona C, González-Neira A, Santisteban P, Robledo M. **Identification of Low Penetrance Genes associated to thyroid cancer susceptibility using a two-step case-control approach.** European Society of Human Genetics Conference. May 23th-26th 2009, Vienna (Austria)

3: **Landa I**, Montero-Conde C, Malanga D, De Gisi S, Letón R, De Marco C, Viglietto G, Robledo M. **A case-control study of CDKN1B (p27) polymorphisms in thyroid cancer identifies a significant SNP with an effect at the transcriptional level.** 33rd Annual Meeting of the European Thyroid Association. September 20th-24th 2008, Thessaloniki (Greece). **Young Investigator Award 2008.*

4: **Landa I.** **Identificación de genes de baja penetrancia asociados al desarrollo de cáncer de tiroides de estirpe folicular mediante genotipado de SNPs.** 1ª Reunión Anual del Centro de Investigación en Red de Enfermedades Raras-CIBERER. 12-13 Noviembre 2007, Barcelona (España).

Posters:

1: **Landa I**, Boullosa C, Inglada-Pérez L, Schiavi F, De La Vega L, Leandro-García LJ, De Cubas A, Maliszewska A, Letón R, Marcos R, Opocher G, Rico D, Valencia A, and Robledo M. **An epistatic study of papillary thyroid cancer: a more comprehensive view of its genetic basis.** 14th International Thyroid Congress. September, 11th-16th 2010, Paris (France).

2: **Landa I**, Ruiz-Llorente S, Montero-Conde C, Inglada-Pérez L, Schiavi F, Leskelä S, Pita G, Milne R, Maravall J, Ramos I, Andía V, Rodríguez-Poyo P, Jara-Albarrán A, Meoro A, del Peso C, Arribas L, Iglesias P, Caballero J, Serrano J, Picó A, Pomares F, Giménez G, López-Mondéjar P, Castello R, Merante-Boschin I, Pelizzo MR, Mauricio D, Opocher G, Rodríguez-Antona C, González-Neira A, Matías-Guiu X, Santisteban P, Robledo M. **La variante rs1867277 en el gen FOXE1 confiere susceptibilidad a desarrollar cáncer de**

tiroides mediante el reclutamiento diferencial de los factores de transcripción USF1/USF2. 52º Congreso Nacional de la Sociedad Española de Endocrinología y Nutrición (SEEN). 26-28 Mayo, Salamanca (España). ***Premio Fundación SEEN en Patología Tiroidea.**

3: **Landa I**, Montero-Conde C, Milne R, Cascón A, Rodríguez-Antona C, Letón R, Alonso MR, Leskelä S, López-Jiménez E, Leandro LJ, Pita G, Ruiz-Llorente S, Santisteban P, González-Neira A, Robledo, 'PTC Clinical Group'. **Identification of Low Penetrance Genes associated to Papillary Thyroid Carcinoma by means of High Throughput Genotyping of Single Nucleotide Polymorphisms (SNPs).** 32nd Annual Meeting of the European Thyroid Association. September 1st-5th 2007, Leipzig (Germany).

Además, el trabajo de esta Tesis ha dado lugar a las siguientes **publicaciones:**

1: **Landa I**, Montero-Conde C, Malanga D, De Gisi S, Pita G, Leandro-García LJ, Inglada-Pérez L, Letón R, De Marco C, Rodríguez-Antona C, Viglietto G, Robledo M. **Allelic variant at -79 (C>T) in CDKN1B (p27^{Kip1}) confers an increased risk of thyroid cancer and alters mRNA levels.** Endocr Relat Cancer. 2010 Jun 1;17(2):317-28. Print 2010 Jun. PubMed PMID: 20075119.

2: **Landa I**, Ruiz-Llorente S, Montero-Conde C, Inglada-Pérez L, Schiavi F, Leskelä S, Pita G, Milne R, Maravall J, Ramos I, Andía V, Rodríguez-Poyo P, Jara-Albarrán A, Meoro A, del Peso C, Arribas L, Iglesias P, Caballero J, Serrano J, Picó A, Pomares F, Giménez G, López-Mondéjar P, Castello R, Merante-Boschin I, Pelizzo MR, Mauricio D, Opocher G, Rodríguez-Antona C, González-Neira A, Matías-Guiu X, Santisteban P, Robledo M. **The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors.** PLoS Genet. 2009 Sep;5(9):e1000637. Epub 2009 Sep 4. PubMed PMID: 19730683; PubMed Central PMCID: PMC2727793.

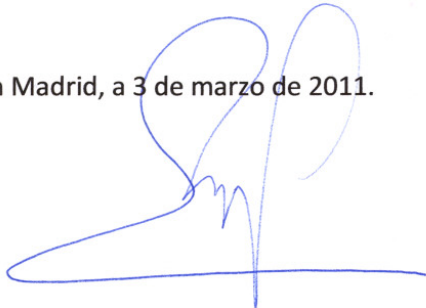
3: Landa I & Robledo M. **Association studies in thyroid cancer susceptibility. Are we on the right track?** J Mol Endocrinol. 2011 [in press].

Iñigo Landa López ha desarrollado su actividad con gran responsabilidad, eficacia y entusiasmo. Su inquietud y su incansable búsqueda de conocimiento ha propiciado que varios de los trabajos hayan recibido **premios en foros internacionales y nacionales:**

- 1- Premio al mejor trabajo de Oncogenética Molecular “Prof. Manuel González Barón” (octubre de 2010).
- 2- Premio Fundación SEEN en Patología Tiroidea (mayo de 2010).
- 3- *Young Investigator Award*, concedido por la *European Thyroid Association* (septiembre de 2008).

Estos premios dan cuenta una vez más de la calidad del trabajo que Iñigo Landa ha venido desarrollando en estos años. Por todo ello, y habiendo revisado el presente trabajo, considero que éste tiene la debida calidad para su presentación y defensa.

En Madrid, a 3 de marzo de 2011.

A handwritten signature in blue ink, consisting of a large, stylized 'M' or 'R' shape with a horizontal line extending to the left and a vertical line extending downwards.

Vº Bº de la Directora de la Tesis

Mercedes Robledo

La presente Tesis Doctoral se realizó en el Centro Nacional de Investigaciones Oncológicas (CNIO) de Madrid entre los años 2006 y 2011 bajo la supervisión de la
Dra. Mercedes Robledo

Las siguientes becas, ayudas y proyectos científicos han permitido la realización de esta Tesis Doctoral:

- Ayuda Predoctoral de Formación en Investigación en Salud (FIS) FI07/00326 del Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación.
- Beca Predoctoral del Programa Nacional de Formación de Profesorado Universitario (FPU) AP2006-01218, Ministerio de Educación.
- Proyecto del Fondo de Investigaciones Sanitarias PI080883 del Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación.

**Carcinoma de tiroides de estirpe folicular:
identificación y caracterización funcional
de genes de baja penetrancia**

A mis padres, por enseñarme a pensar

A Nerea, por estar siempre ahí

A Steph, por demasiadas cosas

Casi cinco años, aunque pasan volando, dan para mucho. En mi caso, para mucho que agradecer, que es lo que pretendo a continuación. Creo que soy un afortunado, un tipo con suerte, o al menos así lo siento. Llegué a Madrid gracias a Meme, por lo que pienso que debe ser la primera a la que dar las gracias. Su confianza desde el primer momento, su disponibilidad, su complicidad durante este tiempo, su cariño, ha hecho realmente fácil trabajar aquí. Sé que será imposible encontrar a alguien así en mi futuro profesional, y eso es difícil de asumir.

En segundo lugar, creo que tuve mucha suerte al caer en un grupo como el de Cáncer Endocrino Hereditario del CNIO. Aquí he podido compartir poyata, mesa, barra, con personas que me han ayudado mucho en mi formación, pero sobre todo, con las que ha sido un gusto pasar tantas horas. Levantarme cada mañana para venir al tajo ha resultado casi indoloro. Si encuentras un trabajo que te gusta, y encima, te rodeas de estas personas, no se puede pedir más.

Gran parte de esta tesis no hubiese sido posible sin el trabajo, el cariño, los buenos consejos de Cristina Montero, que me enseñó un montón de cosas en los dos años y medio que compartimos en el laboratorio, y sin el gran trabajo, generosidad y sintonía con Sergio Ruiz. No me olvido de Leipzig-Berlín, Tesalónica, etc. Sois tan brillantes como humildes, y tan buena gente que sólo podíais acabar juntos. Me habéis tratado tan bien que corréis un serio peligro de que os persiga por Manhattan para seguir parasitándoos un poco más...

Siguiendo con "los endocrinos" (ese grupo carismático como pocos), muchas gracias, Alberto, por tu ayuda y consejos durante este tiempo, por las múltiples conversaciones gafapásticas de cine, música, vocablos extraños, por tu cercanía, por aguantar a los niñatos de veintipocos que empezamos la tesis. A Rocío, porque compartir mesa contigo ("estar a mano galleta", en letonés) ha sido un gusto, una carcajada constante. Gracias por tu cariño, por ser testigo, según tú, de "esos años tan importantes", por modernear conmigo y con tu Mark en los conciertos (y perdón por mis gruñidos de vez en cuando). A Cristina Rodríguez, por toda tu ayuda siempre que la he necesitado, por ser tan buena persona, y la que siempre se tomaba otra con "los jóvenes", como en Las Palmas. A Javi, por ser tan listo, tan rápido, por tener siempre algo que decir, por los servicios logísticos prestados, porque me encanta hablar contigo, por decir lo que piensas, porque me caes muy bien y quiero que esto dure mucho tiempo (así que no te rajes del plan NYC, ¿eh?). A Leti y a Álvaro, por su inestimable ayuda con el *high-throughput genotyping* manual. A Lucía, por saber mucho más de Estadística de lo que yo nunca siquiera me plantearé. A Iñaki (no

sinónimo de Iñigo, por si quedan dudas), para que sigas igual de sensato a pesar del *Nature* en ciernes.

A los grupos hermanos del Programa de Genética del Cáncer Humano, donde, por cierto, se respira mejor que en cualquier otro sitio del CNIO (gracias, JB, por fomentar esto). A Bárbara, Laura, Eva, Javi G, Miljana, Ali, por los buenos ratos. A Fátima, porque eres genial. A María, por enseñarme a resumir y a expresar mis ideas en pocas palabras, aunque, como podrás comprobar, aún tienes trabajo por hacer. A Ana Osorio (la investigadora con más clase del lugar), porque eres de las mejores personas que he conocido y te mereces lo mejor. A Samuel, por ser tan majo, sincero, generoso. A Alba y Ana, de Cito, por su empeño en que los seminarios mejoren. A Roger, por preocuparte de mis análisis estadísticos y de mi vida en general, porque da gusto hablar contigo de cualquier cosa. A Lara, porque echo de menos que estés por aquí todos los días (siempre falta uno cuando hago el recuento antes de irnos).

Al CEGEN, los que más curran del Programa, por toda la ayuda y por ese pedazo de club de tartas, que tenemos que retomar pronto. En especial a Daniela, mi compatriota de la República Independiente de Malasaña, y futura socia en nuestra empresa de éxito (CoolGenoPop®, OGT o las que vengan); a Guille, por toda su ayuda en el análisis informático (mutilaciones de *P* incluidas) y su disposición a cualquier evento social, sobre todo tercio en mano; y a Charo, por ayudarme tanto.

Me acuerdo también de los que empezaron conmigo, aunque ahora estén más o menos lejos. Si no estoy más en contacto con vosotros es culpa mía. A Magda, por su aprecio y por las conversaciones filosóficas. A Sergio Carilla, por ofrecerme su amistad incondicional. Cómo no, a Susanna y a Loren, con los que conecté casi desde el principio cuando llegué al CNIO, y que han acabado convirtiéndose en mis amigos, aunque ya no nos veamos a diario. Añoro las conversaciones de todo tipo que teníamos, los viajes que hicimos, las noches por ahí... ¿Volverán algún día?

Tampoco quiero olvidarme de quienes han sido parte muy importante en este trabajo. A César Boullosa, por pelearse con datos, algoritmos y versiones, y poner un poco de orden en ese gran lío llamado epistasia. A Pilar Santisteban (y su grupo), por todo el trabajo con los ensayos funcionales, por su cercanía, estima, por convertirse un poco a la SNPología, por ser mi apoyo en los congresos europeos de tiroides.

Yéndome un poco más lejos, a Juan Carlos Alonso, del CNB, y, sobre todo, a José Luis Vizmanos, de la UN, por ser el primero en dejarme enredar en un laboratorio, y por su cercanía y aprecio en años posteriores.

No quiero olvidarme del mundo extra-profesional, que me llena, me equilibra, me ayuda a relativizar, a no volverme loco, a distinguir lo importante de lo que no lo es en absoluto, y al que le regateo demasiado tiempo por culpa de este trabajo.

A Xabi, Iban, Sergio y compañía, que están ahí cada vez que vuelvo, que me recuerdan de dónde vengo, que creo que ahí seguirán por mucho tiempo, aunque “el madrileño” no tenga pinta de volver. Muchas veces me descubro a mí mismo pensando en los buenos momentos que he pasado con vosotros desde que nos conocimos en el instituto, de lo que parece que hace 200 años...

A Luis, por ser mi gran amigo desde la universidad, por tantos fines de semana de cañas, copas, conciertos, conversaciones trascendentales, porque te lo mereces todo. A Angie, Ana, y sobre todo, a Laura, que hicieron que compartir piso fuese muy divertido. A Intza, Adri, Patricia, porque me encanta estar con vosotras.

A Urko, por ser un dictador benévolo, el líder carismático de un grupo que acabó convirtiéndose en mis amigos en Madrid. A Pablo, Lula, Javi (-79TT!), los otros miembros destacados del exclusivo e ilustre Film Nerd Group de cada semana (y van ciento y pico pelis). Creo que es una de las cosas que más echaré de menos cuando emigre, así que habrá que inventar algo para seguir haciéndolo (¿teletransporte?).

A la “familia” americana en Madrid: Marc, Shannon, Julie, Carlos, Aaron, Beth, por muchos buenos ratos, por las *family dinners* del pasado. A Diego, para que vuelva pronto. A la familia americana “de verdad”: MaryAnn y Jerry. Gracias por vuestro cariño, vuestra hospitalidad, por tratarme como uno más desde el principio, por ser parte de “la otra América”, la que merece la pena. A mis tíos y no-tíos: Kike, Isabel, Jesús y Ana, por su cariño desde que no sabía ni hablar, porque sé que se alegran un montón de que me vaya bien, y porque les correspondo demasiado poco.

Y aunque sea una pedantería (que lo es) a todos esos señores viejunos, a los que nunca conoceré personalmente, que ya lo hicieron todo en los 60-70s, pero que me siguen ayudando con su música a pasar largas horas delante de la pantalla del ordenador, y a reconciliarme un poco con el mundo.

Y por último, aunque es lo más importante, al núcleo duro de mi vida. A “mis aitas”, que se han pasado la vida cuidándome, que me enseñaron a pensar por mí mismo, que me han dado todo sin imponerme nada, que respetan y apoyan todas mis decisiones, que me quieren tanto. A Nerea, que siempre ha estado ahí, porque es un apoyo constante, por ser la mejor hermana imaginable. Y, por supuesto, a Steph, esa pequeña guiri de sonrisa ancha, porque además de guapa, inteligente, divertida, brillante, encima quiere compartir todo eso conmigo. Lo que decía: un tipo con suerte.

ÍNDICE GENERAL	21
ÍNDICE DE FIGURAS	27
ÍNDICE DE TABLAS	29
ABREVIATURAS	31
RESUMEN/ABSTRACT	33
INTRODUCCIÓN	39
1. Tiroides	41
2. Cáncer de tiroides	43
2.1. Aspectos generales. Epidemiología, diagnóstico y tratamiento	43
2.2. Clasificación de los carcinomas de tiroides de células foliculares	44
2.2.1. Carcinoma Papilar de Tiroides (CPT)	44
2.2.1.1. Carcinoma Papilar de Tiroides puro o clásico (CPTc)	46
2.2.1.2. Carcinoma Papilar de Tiroides variante folicular (CPTvf)	46
2.2.1.3. Otros subtipos del Carcinoma Papilar de Tiroides	46
2.2.2. Carcinoma Folicular de Tiroides (CFT)	47
2.3. Formas de presentación de los CPT y CFT	49
2.3.1. CPT y CFT asociados a síndromes hereditarios	49
2.3.2. CPT y CFT no sindrómicos	50
3. Las enfermedades complejas o multifactoriales	52
3.1. Definición y aspectos generales	52
3.2. Evolución en los métodos empleados en la identificación de genes de baja penetrancia	53
3.3. Ejemplos de enfermedades complejas	56
3.4. El carcinoma de tiroides de estirpe folicular como enfermedad compleja	57
OBJETIVOS	59
MATERIAL y MÉTODOS	63
1. Obtención de muestras y datos clínicos	65
1.1. Obtención de muestras biológicas de pacientes y controles	65
1.1.1. Serie 1 de casos y controles	65

1.1.2. Serie 2 de casos y controles	66
1.1.3. Serie 3 de casos y controles	67
1.2. Obtención de datos clínicos	68
2. Extracción y procesamiento de ADN	70
2.1. Extracción de ADN genómico a partir de linfocitos de sangre periférica	70
2.1.1. Extracción automática de ADN genómico mediante <i>MagNAPureLC</i>	70
2.1.2. Extracción de ADN genómico con el kit de <i>FlexiGene</i>	71
2.2. Extracción de ADN genómico a partir de saliva	71
2.3. Amplificación del genoma completo mediante GenomiPhi™	73
2.4. Cuantificación del ADN genómico mediante PicoGreen	74
3. Selección de genes y polimorfismos para el estudio de asociación	75
3.1 Selección de genes	75
3.2 Selección de SNPs.	83
3.2.1. Criterios de selección de SNPs relativos a su localización	83
3.2.1.1. SNPs utilizados como marcadores o tagSNPs	83
3.2.1.2. SNPs potencialmente funcionales	84
3.2.2. Criterios adicionales de selección de SNPs	85
4. Estrategias de genotipado de polimorfismos	87
4.1. Plataforma de genotipado Illumina GoldenGate®	87
4.1.1. Descripción de la plataforma	87
4.1.2 Principios metodológicos del sistema Illumina GoldenGate	88
4.1.3. Interpretación de los resultados	91
4.1.4. Protocolo específico de genotipado	92
4.2. Plataforma de genotipado TaqMan®	93
4.2.1. Principios metodológicos del genotipado mediante sondas TaqMan®	93
4.2.2. Protocolo específico de genotipado	94
4.3. Plataforma de genotipado KASPar	96
4.3.1. Principios metodológicos del genotipado mediante sondas KASPar	96

4.3.2. Protocolo específico de genotipado	97
4.3.3. Interpretación de los resultados de genotipado por TaqMan y KASPar	101
4.4. Genotipado por digestión diferencial de fragmentos (RFLP)	102
4.5 Genotipado mediante DHPLC	103
4.6. Genotipado por secuenciación directa	104
5. Ensayos funcionales realizados en linfocitos de voluntarios sanos	106
5.1. Obtención de linfocitos a partir de sangre periférica	106
5.2. Ensayos relativos al ARN mensajero	107
5.2.1. Extracción de ARN	107
5.2.2. Retrotranscripción del ARN a ADN complementario	107
5.2.3. PCR cuantitativa a tiempo real	108
5.3. Ensayos con proteína	110
5.3.1. Extracción de proteínas	110
5.3.2. Cuantificación de proteínas mediante Bradford	110
5.3.3. Inmunoblot	111
6. Ensayos funcionales con líneas celulares	113
6.1. Cultivos celulares	113
6.2. Ensayos de luciferasa	114
6.2.1. Construcción de vectores reporteros	114
6.2.1.1. Construcción del vector pGL3-p27 ^{kip1} _{C/T}	114
6.2.1.2. Construcción del vector pGl3b-FOXE1-283G/A	114
6.2.2. Transfecciones y ensayos de luciferasa	115
6.3. Ensayos de retardo (EMSA)	116
6.3.1. Principios metodológicos de los ensayos de retardo	116
6.3.2. Ensayos de retardo referidos a <i>FOXE1</i>	118
7. Estudios de interacción gen-gen o epistasia	119
8. Análisis estadísticos	121
8.1. Estudio individual de SNPs	121

8.2. Estudio de haplotipos	121
8.3. Análisis estadísticos de los estudios funcionales	122
RESULTADOS	123
1. Resultados referidos al gen <i>CDKN1B</i> (p27^{kip1})	127
1.1. El polimorfismo rs34330 (-79C>T) está significativamente asociado con la variante folicular del carcinoma papilar de tiroides (CPTvf)	127
1.2. La variante -79C>T disminuye la tasa de transcripción de <i>CDKN1B</i> en ensayos de luciferasa	131
1.3. El genotipo de riesgo -79TT se correlaciona con menores niveles de ARN mensajero de <i>CDKN1B</i> en linfocitos	134
1.4. Los portadores del genotipo -79TT muestran una tendencia hacia una menor expresión de proteína p27 en linfocitos	135
2. Resultados del SNP array de la primera fase del estudio de asociación	136
2.1. Resultados globales del genotipado en la serie 1	136
2.1.1. Resultados técnicos del genotipado	136
2.1.2. Resultados globales según las frecuencias alélicas	137
2.2. Estudio individual de polimorfismos	138
2.2.1. Resultados en Carcinoma Papilar de Tiroides (CPT)	139
2.2.2. Resultados en Carcinoma Papilar de Tiroides clásico (CPTc)	141
2.2.3. Resultados en la variante folicular del Carcinoma Papilar de Tiroides, (CPTvf)	142
2.2.4. Resultados en Carcinoma Folicular de Tiroides (CFT)	144
2.2.5. Resultados en carcinomas con patrón de crecimiento folicular (CPTvf y CFT)	144
2.2.6. Resultados en carcinomas agresivos	146
2.3. Estudio de haplotipos	147
2.4. Selección de genes de baja penetrancia candidatos para la validación	149
3. Resultados referidos al gen <i>FOXE1</i>	152
3.1. Varios polimorfismos en el <i>locus</i> de <i>FOXE1</i> están significativamente asociados al desarrollo de carcinoma papilar de tiroides en población española	152
3.2. El análisis de haplotipos identifica una combinación de riesgo significativamente asociada al desarrollo de CPTc	152

3.3. Un análisis detallado del haplotipo de riesgo en CPT identifica al SNP rs1867277 como una variante altamente correlacionada y potencialmente funcional.	156
3.4. La asociación significativa para el rs1867277 de <i>FOXE1</i> en carcinoma papilar de tiroides se valida en la segunda fase del estudio	156
3.5. El factor de transcripción DREAM se une al promotor de <i>FOXE1</i> en la región donde se localiza la variante de riesgo rs1867277	159
3.6. Las proteínas USF se unen específicamente al alelo A del rs1867277	161
3.7. Los factores USF1 y USF2 regulan la actividad transcripcional del promotor de <i>FOXE1</i>	163
3.8. Las isoformas α CREM y CREB activan la transcripción de <i>FOXE1</i> mediante un sitio CRE localizado cerca del rs1867277	164
4. Resultados de validación de variantes seleccionadas del SNP array: segunda fase del estudio de asociación	166
4.1. Resultados técnicos del genotipado en las series 2 y 3	166
4.2. Resultados de validación de polimorfismos	167
4.2.1. Resultados en Carcinoma Papilar de Tiroides (CPT)	167
4.2.2. Resultados en Carcinoma Papilar de Tiroides clásico (CPTc)	170
4.2.3. Resultados en la variante folicular del Carcinoma Papilar de Tiroides, (CPTvf)	170
4.2.4. Resultados en Carcinoma Folicular de Tiroides (CFT)	170
4.2.5. Resultados en carcinomas con patrón de crecimiento folicular (CPTvf y CFT)	170
4.2.6. Resultados en carcinomas agresivos	171
4.2.7. Resultados atendiendo a características clínicas del paciente	171
5. Resultados del estudio de interacción gen-gen	173
DISCUSIÓN	177
1. La variante rs34330 (-79C>T) en el gen <i>CDKN1B</i> (p27) confiere riesgo a desarrollar CPTvf y altera los niveles de ARNm	181
2. La variante rs1867277 en el gen <i>FOXE1</i> confiere susceptibilidad a desarrollar cáncer de tiroides mediante el reclutamiento de los factores de transcripción USF1/USF2	185
3. Otros genes podrían determinar la susceptibilidad a desarrollar carcinomas de tiroides de estirpe folicular	190
3.1. <i>TSHR</i> : gen del receptor de la tirotropina	190

3.2. <i>TG</i> : gen de la tiroglobulina	192
3.3. El gen <i>SMAD7</i>	193
3.4. Otros genes candidatos	194
3.5. El análisis de epistasia revela nuevos genes de susceptibilidad	195
4. Factores genéticos de riesgo al cáncer de tiroides. ¿Hemos avanzado en su conocimiento?	197
CONCLUSIONES	203
BIBLIOGRAFÍA	207
APÉNDICE I. Material suplementario	221
APÉNDICE II. Direcciones web	231
APÉNDICE III. Publicaciones derivadas de la tesis	235
APÉNDICE IV. Otras publicaciones	281

INTRODUCCIÓN

Figura I1. Tiroides, folículos y células foliculares.

Figura I2. Clasificación de los carcinomas bien diferenciados de células foliculares de tiroides.

Figura I3. Contribución multifactorial al desarrollo de enfermedades complejas.

Figura I4. Representación esquemática del tamaño muestral necesario para la detección de los riesgos típicos de enfermedades complejas ($OR \sim 1.5$).

Figura I5. Estrategias de genotipado de polimorfismos en los estudios caso-control.

MATERIAL Y MÉTODOS

Figura M1. Modelo de cuestionario clínico estándar para pacientes con cáncer de tiroides esporádico.

Figura M2. Cuantificación de ADN genómico mediante PicoGreen.

Figura M3. Gen, tipos de SNPs y LD.

Figura M4. Plataforma de genotipado Illumina Sentrix® Array.

Figura M5. Esquema del proceso de genotipado Illumina GoldenGate®.

Figura M6. Análisis de los resultados del genotipado Illumina GoldenGate.

Figura M7. Representación esquemática del genotipado mediante sondas TaqMan®.

Figura M8. Representación esquemática del genotipado mediante el sistema KASPar.

Figura M9. Análisis de los genotipos obtenidos con las plataformas TaqMan o KASPar.

Figura M10. Genotipos representativos del rs36228499, discriminados por RFLP.

Figura M11. Discriminación alélica mediante DHPLC.

Figura M12. Representación esquemática de los resultados de PCR cuantitativa

Figura M13. Esquema de los ensayos de retardo o EMSAs

Figura M14. Análisis de interacción gen-gen o epistasia.

RESULTADOS

Figura R1. Actividad transcripcional de las construcciones controladas por el promotor del gen *CDKN1B*, conteniendo las variantes -79C o -79T, en células HeLa.

Figura R2. Actividad transcripcional de las construcciones controladas por el promotor del gen *CDKN1B*, conteniendo las variantes -79C o -79T, en células de tiroides de rata.

Figura R3. Expresión diferencial de *CDKN1B* en linfocitos humanos, distribuidos de acuerdo al genotipo del polimorfismo -79C>T (rs34330).

Figura R4. Niveles de proteína p27^{kip1} en linfocitos humanos, evaluados mediante inmunoblot, en función del genotipo del polimorfismo -79C>T (rs34330).

Figura R5. Representación gráfica de los resultados globales del genotipado de variantes polimórficas en la serie 1.

Figura R6. Resultados globales de la distribución de frecuencias alélicas entre casos (n=609) *versus* controles (n=504) en función de la frecuencia (MAF) de las variantes estudiadas.

Figura R7. Gráfico Q-Q de la distribución de las frecuencias alélicas de los 690 SNPs estudiados en casos frente a controles.

Figura R8. Manhattan plot de la distribución de frecuencias alélicas entre CPT y controles

Figura R9. Manhattan plot de la distribución de frecuencias alélicas entre CPTc y controles

Figura R10. Manhattan plot de la distribución de frecuencias alélicas entre CPTvf y controles

Figura R11. Representación del patrón de desequilibrio de ligamiento (LD plot) en el *locus* de *FOXE1*, basado en los datos de la población HapMap-CEU.

Figura R12. Análisis de la secuencia de la región del promotor de *FOXE1* donde se encuentra la variante rs1867277.

Figura R13. Ensayos de retardo para evaluar la unión de proteínas nucleares a la región del polimorfismo rs1867277 de *FOXE1*.

Figura R14. Ensayos de retardo para evaluar la unión de proteínas nucleares a la región del polimorfismo rs1867277 de *FOXE1* (II).

Figura R15. Actividad transcripcional del promotor de *FOXE1*.

Figura R16. Tasas de error en el genotipado de los 11 SNPs elegidos en la fase de validación

Figura R17. Red de interacciones gen-gen identificadas en la serie 1.

Figura R18. Resultados de epistasia para las cuatro interacciones validadas.

DISCUSIÓN

Figura D1. Mecanismos de regulación negativa de p27.

Figura D2. Mecanismo de regulación propuesto para el promotor de *FOXE1*, según nuestros resultados y lo descrito en otras publicaciones.

Figura D3. Detalle del *locus* de *TSHR*.

Figura D4. Representación esquemática de la contribución de variantes genéticas al desarrollo de carcinomas de tiroides de estirpe folicular, según el riesgo (OR) atribuible a cada variante y su frecuencia alélica (MAF).

Figura D5. Modelo de heterogeneidad genética y ambiental en la predisposición individual al desarrollo de carcinomas de tiroides de estirpe folicular.

MATERIAL SUPLEMENTARIO

Figura S1. Diseño del estudio de asociación en dos pasos para *FOXE1*.

MATERIAL Y MÉTODOS

Tabla M1. Características principales de los grupos de casos y controles incluidos en el estudio

Tabla M2. Genes seleccionados para el estudio y polimorfismos seleccionados en cada uno de ellos

Tabla M3. Polimorfismos estudiados mediante sondas KASPar

Tabla M4. Polimorfismos estudiados por secuenciación directa

Tabla M5. Secuencias de los oligonucleótidos empleados en PCR cuantitativa

RESULTADOS

Tabla R1. Polimorfismos de *CDKN1B* y muestras genotipadas

Tabla R2. Riesgos a desarrollar cáncer de tiroides en función del genotipo de rs34330 (-79C>T)

Tabla R3. Riesgos estimados para los SNPs de *CDKN1B* en la variante folicular del carcinoma papilar de tiroides (CPTvf)

Tabla R4. Resultados significativos ($P<0.05$) al comparar CPT frente a controles en la serie 1.

Tabla R5. Resultados muy significativos ($P<0.01$) al comparar el subtipo CPT clásico frente a controles en la serie 1.

Tabla R6. Resultados muy significativos ($P<0.01$) al comparar la variante folicular de CPT (CPTvf) frente a controles en la serie 1.

Tabla R7. Resultados muy significativos ($P<0.01$) al comparar casos de carcinoma folicular de tiroides (CFT) frente a controles en la serie 1.

Tabla R8. Resultados muy significativos ($P<0.01$) al comparar los carcinomas con patrón de crecimiento folicular (CPTvf+CFT) frente a controles en la serie 1.

Tabla R9. Resultados muy significativos ($P<0.01$) al comparar carcinomas agresivos frente a controles en la serie 1.

Tabla R10. Resultados significativos ($P<0.05$) de la distribución de haplotipos entre casos y controles de la serie 1.

Tabla R11. Genes de baja penetrancia y polimorfismos asociados consistentemente al desarrollo de carcinomas de tiroides de estirpe folicular en la serie 1.

Tabla R12. Polimorfismos significativos en el *locus* de *FOXE1* al comparar casos de carcinoma papilar de tiroides frente a controles de la serie 1.

Tabla R13. Análisis de haplotipos en el *locus* de *FOXE1* en casos de carcinoma papilar de tiroides clásico (CPTc) frente a controles de la serie 1

Tabla R14. Resultados de asociación combinados para la variante rs1867277 (c.-283G>A) en carcinomas papilares de tiroides frente a controles

Tabla R15. Resultados obtenidos en la fase de validación (series 2+3) para las 11 variantes candidatas.

Tabla R16. Comparación de los resultados obtenidos en las fases de descubrimiento (serie 1) y de validación (series 2+3) para las 11 variantes candidatas.

Tabla R17. Resultados de las cuatro interacciones gen-gen validadas.

MATERIAL SUPLEMENTARIO

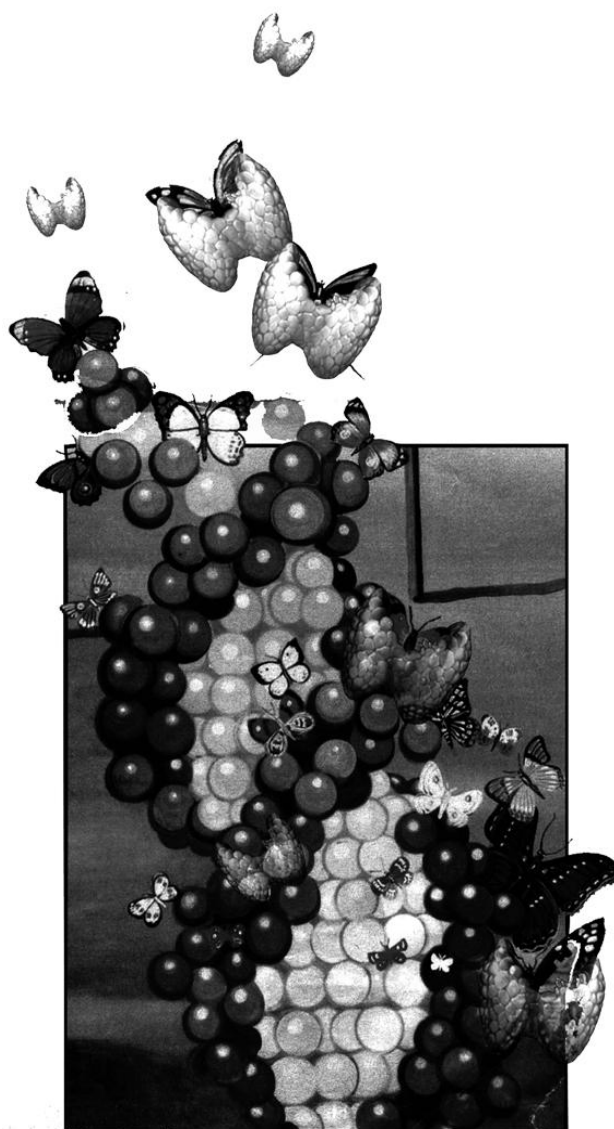
Tabla S1. Listado de SNPs estudiados en la serie 1 del estudio de asociación.

Tabla S2. Desequilibrio de ligamiento (LD) en el *locus* de *FOXE1*.

ADN	Ácido Desoxirribonucleico
ADNc	ADN complementario
ADNg	ADN genómico
AMPc	adenosín monofosfato cíclico
ARN	Ácido Ribonucleido
ARNm	ARN mensajero
<i>αCREM</i>	<i>cAMP responsive element modulator</i>
αTUB	alfa-tubulina
CD/CV	<i>Common Disease / Common Variant</i> (hipótesis)
<i>CDKN1B</i>	<i>cyclin-dependent kinase inhibitor 1B</i>
CFT	Carcinoma Folicular de Tiroides
CMT	Carcinoma Medular de Tiroides
CNMT	Carcinoma No Medular de Tiroides (fundamentalmente CPT y CFT)
CO₂	dióxido de carbono
CPT	Carcinoma Papilar de Tiroides
CPTc	CPT clásico o puro
CPTvf	CPT, variante folicular
CRE	<i>cAMP response element</i>
<i>CREB</i>	<i>cAMP responsive element binding protein</i>
Ct	ciclo umbral (" <i>cycle threshold</i> ") de PCR cuantitativa
DE	desviación estándar
DEPC	dietilpirocarbonato
dNTPs	desoxinucleótidos
DRE	<i>damage response element</i>
<i>DREAM</i>	<i>DRE-antagonist modulator</i>
DTT	ditiotreitól
EDTA	ácido etilendiaminotetracético
EMSA	<i>Electrophoretic Mobility Shift Assay</i>
<i>FOXE1</i>	<i>forkhead box E1</i>
g	gramos [unidad de masa]
<i>g</i>	aceleración de la gravedad=9.8 m/s ²
G₁	fase "gap 1" o de crecimiento en el ciclo celular
GBP	gen de baja penetrancia
GWAS	<i>Genome Wide Association Studies</i>
h	horas [unidad de tiempo]
IC 95%	intervalo de confianza al 95% del <i>odds ratio</i>
Kb	kilobases de ADN
kDa	kilodaltons [unidad de peso molecular de proteínas]
LD	desequilibrio de ligamiento (" <i>Linkage Disequilibrium</i> ")
LR	<i>likelihood ratio</i>
M	molar= mol/l [unidad de concentración]
mA	miliamperios [unidad de intensidad de corriente eléctrica]
MAF	frecuencia del alelo minoritario de un SNP (" <i>Minor Allele Frequency</i> ")
mCi	milicurios [unidad de radiactividad]
MgCl₂	cloruro de magnesio
min	minutos [unidad de tiempo]
ml	mililitros [unidad de volumen]
mM	milimolar= 10 ⁻³ mol/l [unidad de concentración]
μg	microgramos [unidad de masa]
μl	microlitros [unidad de volumen]
μm	micrómetros [unidad de longitud]
NaCl	cloruro sódico
ng	nanogramos [unidad de masa]
nm	nanómetros [unidad de longitud]
O₂	oxígeno
°C	grados centígrados [unidad de temperatura]
OR	riesgo relativo (" <i>odds ratio</i> ")
P	valor de probabilidad, significación estadística

PAAF	punción aspiración con aguja fina
pb	pares de bases de ADN
PBS	tampón fosfato salino (<i>Phosphate Buffered Saline</i>)
PCR	Reacción en cadena de la polimerasa (<i>Polymerase Chain Reaction</i>)
PDYN	<i>prodynorphin</i>
rpm	revoluciones por minuto [unidad de frecuencia]
s	segundos [unidad de tiempo]
S	fase de síntesis o de replicación del ADN en el ciclo celular
S/AS	sentido/antisentido, referido a la dirección de lectura del ADN
SNP	Polimorfismo de Nucleótido Único (<i>Single Nucleotide Polymorphism</i>)
TBS	tampón tris salino (<i>Tris Buffered Saline</i>)
Th	temperatura de hibridación o <i>annealing</i> de la PCR
TSH	tirotropina (" <i>thyroid stimulating hormone</i> ")
TxNxMx	código de clasificación de tumores, siendo T=tamaño del tumor, N=presencia [1]/ausencia [0] de metástasis en nódulos linfáticos locales, y M=presencia[1] /ausencia [0] de metástasis a distancia
U	unidades de enzima
UR	unidades relativas
US	ultrasonografía
USF1	upstream transcription factor 1
USF2	upstream transcription factor 2
V	voltios [unidad de voltaje]

RESUMEN/ABSTRACT



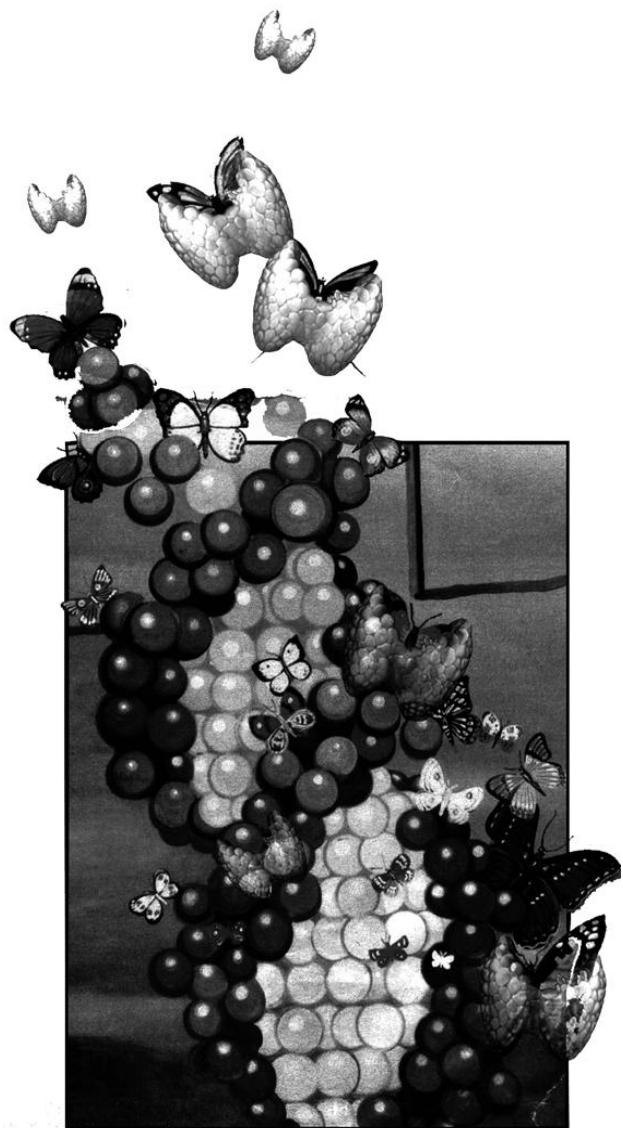
Los carcinomas de tiroides de células foliculares se comportan como una enfermedad compleja o multifactorial, en la que múltiples variantes, localizadas en genes de baja penetrancia (GBP) interactúan entre sí y con el ambiente, determinando la susceptibilidad individual. En este contexto, sólo los estudios de asociación tienen la potencia adecuada para detectar GBP. En este trabajo, hemos realizado un estudio caso-control en dos pasos, empleando una aproximación de genes candidatos. En la primera fase del estudio o serie de descubrimiento, que incluyó más de 600 casos y más de 500 controles, pareados por género y edad, estudiamos un grupo de genes cuidadosamente seleccionados por su papel biológico y/o expresión diferencial en estos tumores. Dichos genes se representaron con variantes polimórficas, tanto tagSNPs como polimorfismos potencialmente funcionales. Además, durante el desarrollo de esta tesis, recopilamos una segunda serie de muestras, que contó con 1020 casos y 1040 controles, y fue empleada como fase de validación de las variantes más significativas de la primera fase. Esta aproximación nos permitió identificar varios GBP relacionados con la susceptibilidad al cáncer de tiroides. En primer lugar, identificamos el polimorfismo rs34330 (c.-79C>T) del gen *CDKN1B* (p27) como un factor de riesgo a desarrollar la variante folicular del carcinoma papilar de tiroides (CPTvf) en población española (OR=2.12; IC 95%=1.09-4.15; $P=0.023$). Además, demostramos que los portadores del genotipo de riesgo rs34330TT muestran una tasa de transcripción menor de *CDKN1B*, lo que supone un mecanismo alternativo de regulación negativa de p27. En segundo lugar, demostramos que la variante rs1867277 (c.-283G>A) del gen *FOXE1* está asociada de forma muy significativa con el desarrollo del carcinoma papilar de tiroides (CPT) en tres poblaciones caucásicas independientes (OR (por alelo)=1.45; IC 95 % =1.30-1.61; $P=4.7 \times 10^{-12}$). El mecanismo subyacente que explicaría la susceptibilidad es un reclutamiento diferencial de los factores de transcripción USF1 y USF2 al promotor de *FOXE1*, que se produce exclusivamente en presencia del alelo de riesgo rs1867277A. Otros *loci* candidatos a actuar como GBPs identificados en nuestro estudio son *TSHR*, *TG* y *SMAD7*, aunque sus variantes causales no pudieron ser determinadas. Por último, mediante un estudio de epistasia o interacción gen-gen, identificamos nuevos GBP potenciales, siendo el gen *PAX8* el más destacado. En resumen, nuestro estudio ha permitido ampliar el conocimiento de los factores genéticos que determinan la susceptibilidad al cáncer de tiroides de estirpe folicular. Además, demuestra que los estudios de genes candidatos, basados en un conocimiento biológico sólido, en un número de muestras amplio y en la adecuada estratificación de pacientes en subgrupos clínicamente homogéneos, constituyen una

aproximación eficaz para la detección de GBP en un modelo de enfermedad compleja como el cáncer de tiroides.

Follicular cell-derived thyroid cancer behaves as a complex or multifactorial disease, in which multiple genetic variants, located on low penetrance genes (LPG), interact with each other and with the environment, thus modulating the individual susceptibility. In this context, only association studies have enough power to detect these LPG. In this work, we performed a two-step, case-control study, adopting a candidate-gene approach. In the first stage of the study, or discovery series, which included more than 600 cases and more than 500 controls matched by age and gender, we studied a group of carefully selected genes due to their biological role and/or differential expression in these tumors. These genes were represented by polymorphic variants, including both tagSNPs and putative functional polymorphisms. In addition, during this thesis, we recruited a second series comprising 1020 cases and 1040 controls, which was used as a validation phase of the top significant associated variants from the first stage. This approach allowed us to identify several LPG related to thyroid cancer susceptibility. First, we identified the rs34330 (c.-79C>T) polymorphism, located on the *CDKN1B* (p27) gene, as a risk factor for developing the follicular variant of papillary thyroid carcinoma (fvPTC) in a Spanish population (OR=2.12; IC 95%=1.09-4.15; $P=0.023$). We further demonstrated that carriers of the rs34330TT risk genotype exhibited a diminished *CDKN1B* transcription rate, thus constituting an alternative p27 downregulation mechanism. Second, we demonstrated that the variant rs1867277 (c.-283G>A) of *FOXE1* gene is highly significantly associated with papillary thyroid cancer (PTC) development in three independent Caucasian populations (OR (per allele)=1.45; IC 95 % =1.30-1.61; $P=4.7 \times 10^{-12}$). We also proposed the differential recruitment of USF1/USF2 transcription factors to *FOXE1* promoter, which takes place exclusively in the presence of rs1867277A risk allele, as the underlying mechanism that could explain the reported susceptibility. In addition, we identified *TSHR*, *TG* and *SMAD7* genes as additional putative LPG, although their respective causal variants could not be determined. Finally, by means of an epistasis or gene-gene interaction study, we identified several new putative LPG, stressing the importance of *PAX8* gene. In summary, our study broadened the knowledge of the genetic factors that determine the susceptibility to follicular cell-derived thyroid cancer development. Moreover, it attests to the efficacy of biologically-oriented, well-powered

and clinically-stratified candidate-gene approaches in the identification of LPG in a complex disease such as thyroid cancer.

INTRODUCCIÓN



1. Tiroides

El tiroides es una glándula endocrina situada en la parte anterior del cuello. Está formada por dos lóbulos, unidos por una porción estrecha de tejido denominado istmo, y en adultos pesa entre 20 y 30 g. La glándula tiroidea dispone de una rica vascularización y de una abundante red de vasos linfáticos (Domarus, et al., 1988).

Desde el punto de vista histológico, la glándula tiroidea está delimitada por una cápsula gruesa de tejido conjuntivo, por debajo de la cual, se dispone una capa interna más fina adherida al tejido glandular. A partir de esta última, se generan pequeños tabiques de tejido conjuntivo que penetran en la glándula y la dividen en lóbulos incompletos. El tiroides está compuesto por acinos o folículos cerrados, de tamaño variable (15-500 μm), que tridimensionalmente forman sacos esféricos, dejando una cavidad central que contiene un material viscoso denominado coloide, compuesto fundamentalmente de la proteína tiroglobulina. Los folículos están revestidos de células cúbicas regulares con una altura de 14 a 15 μm , llamadas células foliculares del tiroides o tirocitos. Además, esparcidas por el tejido intersticial o dentro de los folículos se encuentran otro tipo de células denominadas células parafoliculares o células C (figura I1) (Robbins, et al., 1988; Young, et al., 2000).

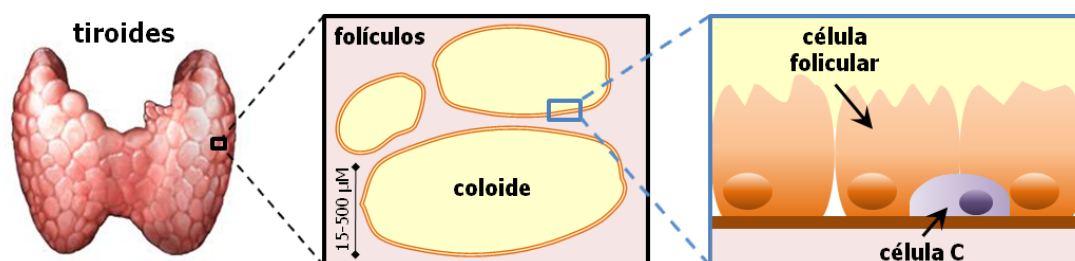


Figura I1. Tiroides, folículos y células foliculares. Representación esquemática de los folículos tiroideos, que son estructuras cerradas con una sustancia coloidal en su interior, formada principalmente por tiroglobulina. Los folículos están constituidos por un epitelio alineado de células polarizadas, conocidas como células foliculares de tiroides o tirocitos, y que constituyen la mayoría del contenido celular tiroideo. Entre ellas, aparecen esporádicamente las células parafoliculares o células C.

Las células foliculares constituyen el 98% del contenido celular tiroideo. Son células epiteliales de origen endodérmico, que se caracterizan por estar polarizadas (disponen de membrana basal y apical), y forman los folículos tiroideos. La función fisiológica de las células foliculares del tiroides consiste en la metabolización del yodo y

la consiguiente biosíntesis y secreción de las hormonas tiroideas: la tiroxina (T4) y la triyodotironina (T3). Estas hormonas participan en multitud de procesos metabólicos, entre los que destaca su capacidad para aumentar la tasa metabólica basal, intervenir en la síntesis proteica y aumentar la sensibilidad a las catecolaminas, además de ser fundamentales para el desarrollo y el crecimiento en la etapa fetal y neonatal (Domarus, et al., 1988; Robbins, et al., 1988).

Por otro lado, las células parafoliculares, o células C, son células epiteliales originadas a partir de la cresta neural (ectodermo), y constituyen una pequeña parte del tejido tiroideo. Carecen de polaridad y están especializadas en la producción de la hormona calcitonina, que participa en el metabolismo del calcio.

2. Cáncer de tiroides

2.1. Aspectos generales. Epidemiología, diagnóstico y tratamiento

El cáncer de tiroides representa el 1 % de todas las neoplasias, con una incidencia anual estimada de 212,000 casos en todo el mundo (Ferlay, et al., 2010). Este número ha ido creciendo rápidamente en los últimos años (Liu, et al., 2001), probablemente debido a las mejoras en las técnicas de diagnóstico (Davies and Welch, 2006), y por el efecto de factores ambientales (Chen, et al., 2009).

El “cáncer de tiroides” es un término general que incluye dos grupos principales de neoplasias, dependiendo del tipo celular afectado por la transformación maligna. Las neoplasias tiroideas más comunes se originan a partir de las células foliculares o tirocitos, el tipo celular responsable de la producción de hormonas tiroideas (DeLellis, 2004). Entre ellas, la gran mayoría pertenecen al grupo de los carcinomas bien diferenciados, siendo los Carcinomas Papilares de Tiroides (CPT) y los Carcinomas Foliculares de Tiroides (CFT) los subtipos más comunes, ya que representan el 85 % y el 10 % de los casos respectivamente (Kondo, et al., 2006). Estos carcinomas presentan en general un buen pronóstico, siendo el número de pacientes que mueren a causa de la enfermedad relativamente pequeño (Ferlay, et al., 2010). Por el contrario, los carcinomas pobremente diferenciados y anaplásicos, que representan un 2% de los carcinomas derivados de células foliculares, se asocian con muy mal pronóstico. De hecho, es una de las neoplasias humanas más agresivas, que conduce a la muerte del paciente en pocos meses. Los carcinomas de tiroides de estirpe folicular suelen denominarse genéricamente como Carcinomas No Medulares de Tiroides (CNMT).

El segundo grupo de tumores tiroideos se origina a partir de las células parafoliculares o células C, constituyendo los llamados Carcinomas Medulares de Tiroides (CMT), que representan el 3 % de los casos (Randolph and Maniar, 2000). Los tumores originados a partir de estas células no son objeto de esta tesis, por lo que no serán considerados de aquí en adelante.

El diagnóstico del CNMT se realiza, en primer lugar, mediante una técnica de imagen conocida como ultrasonografía (US), que permite la detección de nódulos tiroideos. Dado que la presencia de nódulos es bastante frecuente en la población general (4-50 %), y sólo un 5 % de ellos son malignos, hacen falta criterios que los

distingan, siendo los principales la vascularización intra-nodal, la irregularidad de los bordes, el aspecto sólido y la presencia de calcificaciones (Pacini, et al., 2010; Papini, et al., 2002). En aquellos casos en los que la exploración mediante US apunta a un riesgo de malignidad, se realizan estudios de punción y aspiración con aguja fina (PAAF). Esta técnica consiste en la obtención de material de los nódulos tiroideos, cuya morfología es luego evaluada por patólogos en extensiones citológicas, si bien no existen marcadores moleculares idóneos (Pacini, et al., 2010).

Para el tratamiento del CNMT, el procedimiento estándar consiste en una tiroidectomía total o casi total, con resección de los ganglios afectados, si los hubiera. La cirugía se complementa normalmente con la administración posterior de dosis terapéuticas de yodo radiactivo (^{131}I) (≥ 100 mCi), que debido a la capacidad del tiroides para captarlo, permite la ablación del tejido tiroideo remanente, y de los potenciales focos microscópicos de tumor residual (Mazzaferri and Kloos, 2001; Pacini, et al., 2010).

A pesar del buen pronóstico que presentan los carcinomas diferenciados de tiroides, entre el 20 y el 40 % de los pacientes desarrollarán recidivas, por lo que es necesario realizar un seguimiento post-operatorio (Schlumberger, 1998), consistente en la palpación del cuello y ganglios, rastreos corporales totales con radioyodo, exploraciones morfológicas de la región cervical, y determinación de los niveles de tiroglobulina sérica (Pacini, et al., 2010).

2.2. Clasificación de los carcinomas de tiroides de células foliculares

La gran mayoría (>95%) de los carcinomas de tiroides de células foliculares, o CNMT, pertenecen a la categoría de los carcinomas bien diferenciados. Las clases principales de este grupo se describen a continuación y en la figura I2.

2.2.1. Carcinoma Papilar de Tiroides (CPT)

El Carcinoma Papilar de Tiroides (CPT) se define como un tumor epitelial maligno que muestra evidencia de células foliculares diferenciadas, tales como la polaridad y la producción de tiroglobulina, y presenta características nucleares distintivas. Entre estas últimas, destacan la presencia de núcleos alargados e

irregulares, con cromatina de aspecto polvoriento, y nucleolos pequeños (DeLellis, 2004).

El CPT es el grupo más frecuente de carcinomas de células foliculares de tiroides, representando un 85 % de los casos. Su incidencia es notablemente más alta en mujeres que en hombres, con un cociente mujer:hombre en torno a 3:1, lo que sugiere un papel de factores hormonales desconocidos en esta enfermedad. Sin embargo, el único factor ambiental bien establecido relacionado con CPT es la exposición a radiación ionizante (Williams, 2009), y su incidencia es especialmente elevada en países desarrollados. El CPT metastatiza preferentemente vía linfática, favorecido en parte por la abundante red de vasos linfáticos que irrigan los lóbulos tiroideos, de manera que la presencia de invasión en los ganglios cervicales en el momento del diagnóstico es bastante común (Schlumberger, 1998).

Desde el punto de vista molecular, se han identificado varias alteraciones somáticas en tumores papilares, todas ellas implicadas en la activación de la vía de las MAP kinasas: reordenamientos de los genes *RET* y *NTRK1*, y mutaciones puntuales en *BRAF* (Kondo, et al., 2006).

Los reordenamientos del proto-oncogén *RET* con varios genes son bastante frecuentes (13-43 % de los CPT), y llevan a la generación de una proteína quimérica, conocida como RET/PTC, que está constitutivamente activada (Nikiforov, 2002; Santoro, et al., 1992). Aunque existen más de 15 genes que se reordenan con *RET*, los más comunes son *CCDC6* ("RET/PTC1") (Grieco, et al., 1994), *PRKAR1A* ("RET/PTC2") (Sozzi, et al., 1994) y *NCOA4* ("RET/PTC3") (Santoro, et al., 1994). Los reordenamientos RET/PTC son especialmente frecuentes en individuos jóvenes y en casos asociados a la exposición a radiación ionizante, como los relacionados con el accidente nuclear de Chernobyl (50-90 %) (Bongarzone, et al., 1996; Bounacer, et al., 1997; Rabes, et al., 2000). Análogamente a lo descrito para *RET*, se han identificado reordenamientos que fusionan el gen *NTRK1* con otros genes (Beimfohr, et al., 1999; Greco, et al., 1992), aunque su prevalencia es bastante menor (5-13 % de los CPT).

Por otra parte, las mutaciones puntuales en *BRAF*, son un evento muy frecuente (29-69 %) y exclusivo de CPT (Davies, et al., 2002; Xing, 2005). La mutación recurrente en *BRAF*, p.Val600Glu, que implica al mismo residuo comprometido en otros carcinomas como colon y melanoma, afecta al dominio tirosín-kinasa de la proteína, y conduce a la activación constitutiva de la ruta de las MAP kinasas.

El CPT es una neoplasia bastante heterogénea, e incluye más de 12 subtipos histológicos, distinguibles no sólo desde el punto de vista histopatológico, sino también en cuanto a características clínicas (figura I2).

2.2.1.1. Carcinoma Papilar de Tiroides puro o clásico (CPTc)

El CPT convencional, puro o clásico (CPTc) es un tumor de crecimiento lento, bien diferenciado y frecuentemente no encapsulado. El rasgo más característico es la presencia de papilas, formadas por un eje conjuntivo revestido por células epiteliales cúbicas con núcleos alargados y de borde irregular, muchos sin cromatina visible (Domarus, et al., 1988). Los CPTc representan aproximadamente la mitad de todos los CPT.

2.2.1.2. Carcinoma Papilar de Tiroides, variante folicular (CPTvf)

La variante folicular del CPT (CPTvf) presenta características intermedias entre un CPT y un CFT. Así, y a diferencia de los CPTc, forma estructuras foliculares en lugar de papilas, pero mantiene las peculiaridades nucleares propias de un CPT convencional. La inclusión de los CPTvf en uno u otro grupo se ha discutido durante mucho tiempo, y se han realizado varios estudios para definir su comportamiento clínico (Liu, et al., 2006; Zidan, et al., 2003). De todas formas, parece claro que los CPTvf presentan características peculiares, al menos histológicas, y deben ser considerados como una entidad particular, distinta del resto de los CPT (Salajegheh, et al., 2008), y constituyendo aproximadamente una cuarta parte de éstos.

2.2.1.3. Otros subtipos del Carcinoma Papilar de Tiroides

Como se ha comentado antes, la heterogeneidad es una importante característica de los CPT, en el que se distinguen muchos otros subtipos, bien definidos histológicamente (DeLellis, 2004). Sin embargo, su baja frecuencia limita la inclusión de estos carcinomas en estudios, especialmente aquéllos que requieren un amplio número de muestras, como los estudios de asociación. Cabe destacar la "variante esclerosante difusa del CPT", caracterizada por la presencia de células escamosas y

abundantes metástasis locales al diagnóstico, aunque su pronóstico es similar a los CPTc (Fujimoto, et al., 1990). Por el contrario, los "CPT de células altas", caracterizados por la presencia de células tres veces más altas que anchas, tiene uno de los peores pronósticos dentro de los CPT (Ghossein and Livolsi, 2008).

2.2.2. Carcinoma Folicular de Tiroides (CFT)

El Carcinoma Folicular de Tiroides (CFT) se define como un tumor epitelial maligno con evidencias de células foliculares diferenciadas, pero sin las características nucleares diagnósticas que definen al CPT. Los CFT se presentan habitualmente como tumores sólidos, encapsulados, de forma ovoide, y normalmente más grandes (diámetro > 1 cm) que los CPT. Desde un punto de vista histológico, el CFT es un tumor invasivo, generalmente solitario, con características morfológicas variables, que van desde la formación de estructuras foliculares con coloide en su interior, a patrones de crecimiento sólido o trabecular. Su característica citológica más distintiva es la carencia de la atipia nuclear propia del CPT, lo que además permite diferenciar los CFT de los CPTvf, que también presentan estructuras foliculares (DeLellis, 2004).

El CFT representa aproximadamente el 10 % de los todos los casos de carcinomas de tiroides. Es más frecuente en mujeres, aunque en una proporción menos acusada que en los CPT (cociente mujer:hombre=2:1). El desarrollo de CFT se ha asociado con dietas deficitarias en yodo, siendo así más frecuente en países en vías de desarrollo (Woodruff, et al., 2010). Los CFT presentan peor pronóstico que los CPT, y a diferencia de éstos, se propagan vía hematógena, principalmente a pulmón y hueso. Un 20 % de los CFT presentan metástasis a distancia en el momento del diagnóstico, mientras que la expansión vía linfática es poco común (5 %) (DeLellis, 2004).

A diferencia del CPT, existen sólo dos categorías de CFT, definidas en función del grado de invasión de la cápsula tiroidea: CFT mínimamente invasivo y CFT invasivo. El primero se caracteriza por una invasión focal de la cápsula y/o estructuras vasculares, mientras que el segundo presenta una amplia infiltración en el parénquima tiroideo y/o vascular.

Respecto a las alteraciones somáticas descritas en CFT, las más frecuentes son las mutaciones puntuales activantes de la familia de genes *RAS* (Manenti, et al., 1994),

que aparecen en el 40-53 % de los CFT, y los reordenamientos del gen *PPARG* (25-63 %), especialmente la translocación *PPARG-PAX8* (Kroll, et al., 2000; Nikiforova, et al., 2003). Aunque en un principio estas alteraciones se describieron como exclusivas de CFT, sería más correcto decir que sólo aparecen en carcinomas con patrón de crecimiento folicular, ya que algunos estudios sugieren que son relativamente frecuentes en los CPTvf (Castro, et al., 2006; Zhu, et al., 2003).

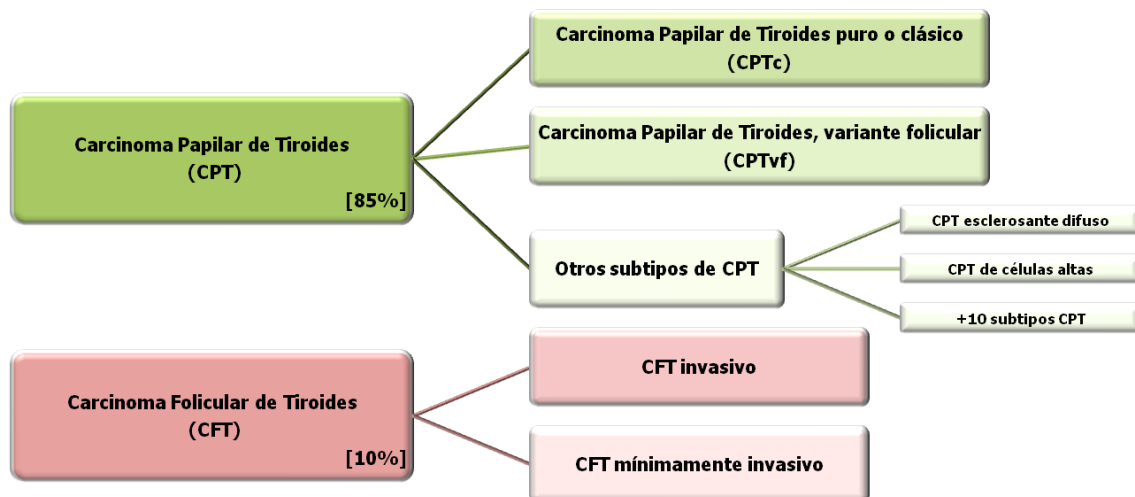


Figura I2. Clasificación de los carcinomas bien diferenciados de células foliculares de tiroides. Otros subtipos minoritarios de CPT, no representados en la figura, son las variantes oncocítica, de células claras, macrofolicular, de células columnares, sólida, cribriforme, con estroma fascicular, con componente focal insular, mucoepidermoide, de células gigantes, los microcarcinomas y los carcinomas papilar y medular combinados.

2.3. Formas de presentación de los CPT y CFT

Los carcinomas de tiroides de estirpe folicular aparecen en su gran mayoría (95 %) como entidades esporádicas. Una pequeña proporción, aproximadamente del 5 %, aparece, bien en el contexto de síndromes hereditarios conocidos, o bien como manifestación única dentro de una agregación familiar, de causas aún no bien caracterizadas.

2.3.1. CPT y CFT asociados a síndromes hereditarios

Una pequeña proporción (2 %) de los CPT y CFT se desarrollan en el contexto de síndromes hereditarios conocidos, en los que los pacientes, además de otros signos principales, muestran una mayor predisposición a desarrollar carcinomas de tiroides de estirpe folicular (Nose, 2008; Vriens, et al., 2009). Estos síndromes se describen a continuación.

La poliposis adenomatosa familiar (FAP), o síndrome de Gardner (MIM #175100), es una enfermedad autosómica dominante caracterizada por la aparición de cientos a miles de pólipos en colon y recto, y causada por mutaciones germinales que inactivan el gen *APC* (Nakamura, et al., 1991). Entre un 1 y un 2 % de los pacientes con FAP desarrollan CPT, con patrón de crecimiento cribriforme (Lee, et al., 2004).

El síndrome de Cowden (MIM #158350) es una enfermedad autosómica dominante caracterizada por la presencia de múltiples hamartomas, y una mayor prevalencia de cáncer de mama, endometrio y tiroides en los pacientes, siendo esta vez el subtipo folicular (CFT) el más frecuente. El gen involucrado en el desarrollo de este síndrome es el supresor tumoral *PTEN* (Nelen, et al., 1996). El 80% de los pacientes con síndrome de Cowden son positivos para mutaciones germinales en *PTEN*, y de ellos, el 10 % desarrolla CFT o su lesión precursora, el adenoma folicular (Harach, et al., 1999).

El síndrome de Werner (MIM #277700) es una enfermedad autosómica recesiva, asociada a mutaciones en el gen *WRN*, y caracterizada principalmente por signos de envejecimiento prematuro y una alta incidencia de neoplasias. Así, en algunos individuos afectados aparecen CFT y CPT.

El complejo de Carney (MIM #160980) es un síndrome multi-neoplásico en el que los afectos pueden desarrollar CPT ó CFT, aunque con una frecuencia baja. El gen principal relacionado con esta enfermedad es *PRKARIA* (Kirschner, et al., 2000), que, como se comentó anteriormente, aparece reordenado con *RET* en un pequeño porcentaje de CPT esporádicos.

Por último, es interesante señalar que, en tumores tiroideos aislados, se han descrito mutaciones somáticas en algunos de los genes asociados a estos síndromes hereditarios, como *APC* o *PTEN* (Hou, et al., 2007; Subramaniam, et al., 2007), aunque aparecen en un porcentaje muy pequeño de los casos esporádicos.

2.3.2. CPT y CFT no sindrómicos

Los carcinomas de tiroides asociados a síndromes hereditarios son una excepción, dado que la mayoría se presentan de forma aislada o en agregación familiar sin seguir un modelo de herencia Mendeliana. En este sentido, está bien establecido que el cáncer de tiroides de estirpe folicular tiene un componente genético muy fuerte, ya que los familiares de primer grado de individuos afectados muestran un riesgo relativo a desarrollar un tumor tiroideo entre 8.6 y 10.3 veces con respecto a la población control, lo que supone el mayor de entre todas las neoplasias que no muestran una herencia Mendeliana clásica (Goldgar, et al., 1994; Hemminki, et al., 2004; Pal, et al., 2001). Además, la contribución genética, conocida como "heredabilidad", se estima que alcanza una proporción del 53 % en el cáncer de tiroides, siendo la única que supera la influencia del ambiente, y duplicando la contribución estimada para otras neoplasias comunes, como el cáncer de mama o el melanoma (Czene, et al., 2002).

Estas observaciones, junto con una evidente agregación familiar en un 5 % de los casos, llevó originalmente a la idea de que los CPT y CFT podrían ser explicados mediante mutaciones en genes de alta penetrancia (Khan, et al., 2010). Varios estudios de ligamiento, utilizando familias no sindrómicas, en la que varios miembros habían desarrollado carcinomas de tiroides de estirpe folicular, exploraron esta posibilidad. Así, se han descrito varios *loci* potenciales de susceptibilidad en las regiones cromosómicas 14q31 (Bignell, et al., 1997), 19p13.2 (Canzian, et al., 1998; Prazeres, et al., 2008), 1q21 (Malchoff, et al., 2000), 2q21 (McKay, et al., 2001), 8p23.1-p22 (Cavaco, et al., 2008), 8q24 (He, et al., 2009) y 6q22 (Suh, et al., 2009).

Estos descubrimientos no han sido replicados de manera uniforme, y es interesante señalar que no se ha identificado ningún gen de alta penetrancia en estas regiones identificadas mediante estudios de ligamiento, probablemente debido a la heterogeneidad de la enfermedad. De hecho, parece que los términos usados comúnmente, tales como "CPT familiar" o "CNMT familiar", incluyen entidades diferentes. Esta idea está apoyada por las características fenotípicas que aparecen en cada familia, y que incluyen varios subtipos histológicos de CPT (CPTvf, tumores oxifílicos, etc.), además de diferentes manifestaciones clínicas adicionales (bocio multinodular, melanoma, etc.). En resumen, no está claro si los carcinomas familiares de estirpe folicular son una entidad aparte (Capezzone, et al., 2008), que podría ser explicada por mutaciones en genes de alta penetrancia aún no identificados, o por el contrario, son el resultado de el efecto combinado de varios *loci* cuya contribución sería entre baja y moderada.

Se considera por tanto al carcinoma de tiroides de estirpe folicular como una enfermedad compleja o multifactorial, causada por la contribución de múltiples genes de baja a moderada penetrancia (GBP) (Adjadj, et al., 2009; Landa and Robledo, 2011; Sturgis and Li, 2009).

3. Las enfermedades complejas o multifactoriales

3.1. Definición y aspectos generales

En los años 90, se enunció formalmente la “hipótesis enfermedad común/variante común” (*Common Disease/Common Variant hypothesis*, o simplemente CD/CV) (Cargill, et al., 1999; Chakravarti, 1999; Lander, 1996). Según ella, el riesgo genético para enfermedades comunes estaría determinado, en muchos casos, por alelos relativamente frecuentes que predispondrían al desarrollo de la enfermedad. Así, las enfermedades complejas o multifactoriales pueden definirse como aquéllas causadas por variantes genéticas comunes, localizadas en genes de baja penetrancia (GBP), que interactúan entre ellas y con factores ambientales, modulando así la susceptibilidad individual. Las variantes polimórficas que determinan esta susceptibilidad pueden alcanzar una alta frecuencia alélica en la población general no afecta, de modo que se espera que su contribución a la enfermedad sea baja o moderada. De esta forma, sólo aquellos individuos que acumulen un número determinado de variantes de riesgo en GBP, y bajo determinadas condiciones ambientales, desarrollarán la enfermedad (figura I3).

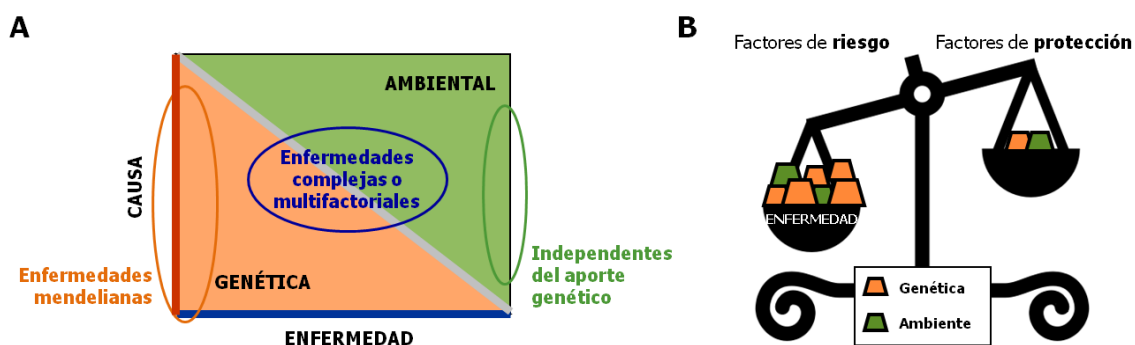


Figura I3. Contribución multifactorial al desarrollo de enfermedades complejas. **A.** Contribución relativa de factores genéticos y ambientales al desarrollo de enfermedades. Las enfermedades de herencia mendeliana se agrupan cerca del eje y, mientras que en las enfermedades multifactoriales, los factores genéticos deben interaccionar con los ambientales para su desarrollo (modificado de Rueff y cols. (Rueff, et al., 2002)). **B.** Representación esquemática de la susceptibilidad individual a desarrollar una enfermedad compleja. Un determinado individuo acumula factores de riesgo y de protección (genéticos y ambientales) a desarrollar una enfermedad, que sólo se manifestará cuando los factores de riesgo superen a los de protección.

Como se detalla a continuación, la estrategia más adecuada para la identificación de GBP es el empleo de los estudios de asociación o caso-control, ya que los estudios de ligamiento no tienen la potencia adecuada para detectar contribuciones genéticas moderadas (Houlston and Peto, 2004; Milne and Benitez, 2008).

3.2. Evolución de los métodos empleados en la identificación de genes de baja penetrancia

La identificación de GBP asociados a enfermedades complejas ha estado supeditada a dos aspectos fundamentales: el reclutamiento de series de muestras lo suficientemente amplias, y el desarrollo de tecnologías de genotipado a gran escala.

La disponibilidad de un tamaño muestral suficiente es fundamental para la detección de factores genéticos de riesgo en enfermedades multifactoriales. El riesgo atribuible a una única variante está entre 1.1 y 1.6 veces para los portadores de los alelos de riesgo (Fletcher and Houlston, 2010). Estos *odds ratios* (OR), junto con la frecuencia con la que las variantes a estudiar aparecen en la población general, determinan la necesidad de un grupo de muestras más o menos amplio para la detección consistente de GBPs. El efecto que tienen estas variables en el diseño de un estudio de asociación se muestra en la figura I4. Según esta simulación, harían falta, por ejemplo, unos 400 casos (y sus correspondientes 400 controles) para detectar un $OR=1.4$, conferido por una variante con una frecuencia en población general del 20 %.

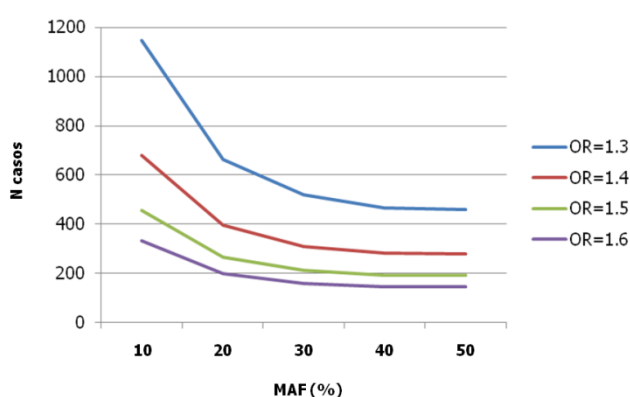


Figura I4. Representación esquemática del tamaño muestral necesario para la detección de los riesgos típicos de enfermedades complejas ($OR \sim 1.5$). El número necesario de pacientes depende de forma importante de la frecuencia del SNP estudiado y del riesgo que éste confiere a desarrollar la enfermedad.

Los cálculos para este gráfico se realizaron con el programa "*PS Power and Sample Size Calculations*", versión 2.1.31 (Dupont and Plummer, 1998), considerando el tamaño muestral necesario para detectar los OR indicados en cada caso, mediante un test de chi-cuadrado, con un nivel de significación $\alpha=0.05$ y una potencia del 80 %. Abreviaturas: N=número de muestras; MAF= *Minor Allele Frequency*; OR= *odds ratio*.

La falta del tamaño muestral necesario suele provocar, en el mejor de los casos, una sobreestimación del riesgo conferido por una variante, y en el peor, la identificación de asociaciones espurias o falsos positivos.

Por otra parte, la generalización de los estudios caso-control sólo ha sido posible con el desarrollo de tecnologías de genotipado masivo. La necesidad de genotipar simultáneamente muchas variantes genéticas en un gran número de muestras llevó al desarrollo de esta tecnología, representada fundamentalmente por los *arrays* de SNPs.

Existen dos estrategias, representadas en la figura I5, para la identificación de variantes de susceptibilidad al desarrollo de una enfermedad compleja. La primera que se empleó, y también la más evidente, es una estrategia directa, que consiste en buscar la variante causal de una enfermedad genotipando aquéllas variantes con algún tipo de posible papel funcional: SNPs que producen cambios de aminoácidos en la proteína resultante, o polimorfismos que modulan la expresión del gen o su procesamiento transcripcional, condicionando la unión de factores de transcripción o reguladores de *splicing*, respectivamente. Esta estrategia, tiene la ventaja de considerar variantes que son interesantes *per se* para explicar el mecanismo de susceptibilidad implicado. Sin embargo, tiene la enorme desventaja de limitarse a estudiar variantes funcionales, que en muchos casos no suelen estar identificadas *a priori*, con la consiguiente pérdida de efectividad.

La aproximación alternativa, conocida como estrategia indirecta, consiste en el aprovechamiento de los datos de desequilibrio de ligamiento (*Linkage Disequilibrium*, LD) disponibles gracias al proyecto HapMap (Consortium, 2003). Así, el método consiste en el genotipado de un número determinado de SNPs marcadores o tagSNPs, que aunque no son funcionales, capturan la variación de un determinado *locus*, incluidas las posibles variantes funcionales contenidas en él. La gran ventaja de esta estrategia es que es una aproximación agnóstica que permite la identificación de GBPs y variantes desconocidas en el momento del estudio, con las que los SNPs utilizados estarán en desequilibrio de ligamiento. Su mayor limitación es la dificultad para interpretar los resultados de asociación, y la posterior búsqueda de la variante causal.

Dado que ninguna de las dos estrategias parece óptima para todos los planteamientos posibles, una aproximación mixta aparece como la más adecuada.

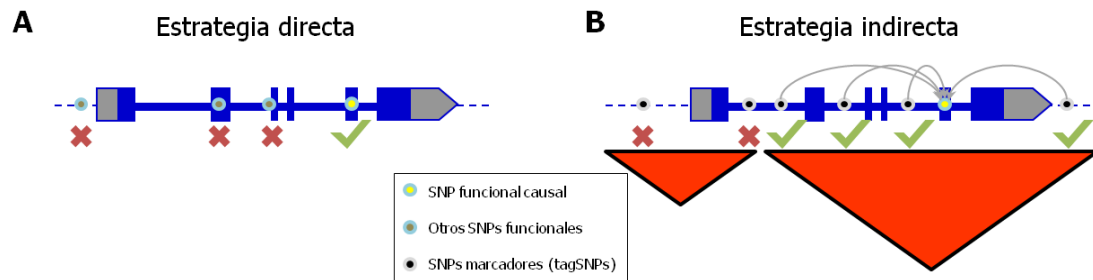


Figura 15. Estrategias de genotipado de polimorfismos en los estudios caso-control. Los exones de un gen hipotético se muestran en azul y gris, los SNPs están representados por círculos, y el LD aparece en rojo. Las cruces rojas y las señales verdes indican resultados de asociación negativos y positivos, respectivamente. **A.** Estrategia directa. Genotipado de variantes potencialmente funcionales: SNPs que producen cambios de aminoácidos o intervienen en la unión de factores de transcripción o reguladores de *splicing*; **B.** Estrategia indirecta. Genotipado de SNPs marcadores o tagSNPs que, a través del LD capturan la variación de una región, marcando la variante causal de la enfermedad.

Por otra parte, surge la pregunta de qué regiones o *loci* estudiar para la identificación de GBP en una enfermedad determinada. Para esto, también existen dos posibles estrategias: los estudios de genes candidatos y los estudios del genoma completo (*Genome Wide Association Studies* o GWAS). La aproximación de genes candidatos tiene la gran ventaja de dotar de un sentido biológico y una interpretación bastante evidente, además de ser un método compatible con las estrategias directas e indirectas. Sin embargo, requiere un detallado conocimiento previo sobre qué papel, mecanismos, expresión, etc., tienen los genes elegidos en el tejido de interés, lo cual requiere estudios preliminares y limita la identificación de genes sin papel conocido.

Por otra parte, los GWAS, se limitan al estudio de una lista de tagSNPs estándar, diseñados por las empresas que suministran la tecnología de genotipado, que capturan la mayor parte de la variación polimórfica del genoma humano. Sus ventajas e inconvenientes son similares a las descritas para la estrategia indirecta: son útiles cuando no se dispone de conocimiento biológico previo sobre la enfermedad, pero fallan en la identificación de variantes causales y en la interpretación biológica (Katsanis, 2009). Además, hay que añadir, a la vista de las asociaciones identificadas mediante GWAS, que parecen funcionar peor para la identificación de variantes menos frecuentes y/o riesgos bajos (Galvan, et al., 2010; Ioannidis, et al., 2010).

En resumen, los métodos y estrategias mencionadas en este apartado, con sus ventajas e inconvenientes, coexisten en la actualidad como aproximaciones alternativas

para la identificación de variantes de riesgo en GBP. Aunque en primera instancia la tendencia general fue hacia una progresiva sustitución de los estudios de genes candidatos por GWAS, la necesidad de encontrar interacciones entre genes, normalmente implicados en la misma ruta metabólica, ha invertido parcialmente esta tendencia, de forma que no se puede decir, en general, que una técnica haya prevalecido sobre otra.

3.3. Ejemplos de enfermedades complejas

En la última década, gracias a la generalización de los sistemas de genotipado masivo, se han identificado asociaciones entre multitud de rasgos fenotípicos y variantes localizadas en GBP. Tal es el caso del color de ojos, el color de pelo o la talla. De forma similar, se han identificado GBP en bastantes enfermedades, consideradas poligénicas o multifactoriales (un amplio catálogo puede consultarse en <http://www.genome.gov/GWASudies/>). Algunos ejemplos bien caracterizados son ciertas enfermedades metabólicas, como las diabetes tipo 1 y 2, enfermedades neurodegenerativas como el Alzheimer o el desorden bipolar, autoinmunes como la enfermedad de Crohn o la artritis reumatoide, y cardiovasculares, tales como el infarto de miocardio o la aterosclerosis (Hindorff, et al., 2009; Manolio, et al., 2009).

Además, existen muchos ejemplos de GBP identificados en varias neoplasias. Las más estudiadas son las variantes de riesgo a desarrollar cáncer de mama, colon y próstata, pero también hay ejemplos consistentes en cáncer de vejiga, pulmón, páncreas, ovario, testículo, melanoma, glioma, neuroblastoma y leucemias (Dong, et al., 2008; Fletcher and Houlston, 2010).

3.4. El carcinoma de tiroides de estirpe folicular como enfermedad compleja

Como se ha mencionado anteriormente, el 95 % de los carcinomas de tiroides de células foliculares se comportan como entidades esporádicas, cuyas causas genéticas apenas son conocidas. Sin embargo, es un hecho que, aparte de los casos claramente atribuibles a efectos ambientales (por ejemplo, los relacionados con el accidente nuclear de Chernobyl en 1986), no se conoce por qué ciertos individuos desarrollan carcinomas de tiroides y otros no. Es lógico pensar, teniendo en cuenta la ya mencionada influencia genética sobre esta enfermedad, que ciertos individuos sean más susceptibles al desarrollo de CPT y CFT porque acumulan variantes polimórficas de riesgo en determinados GBP.

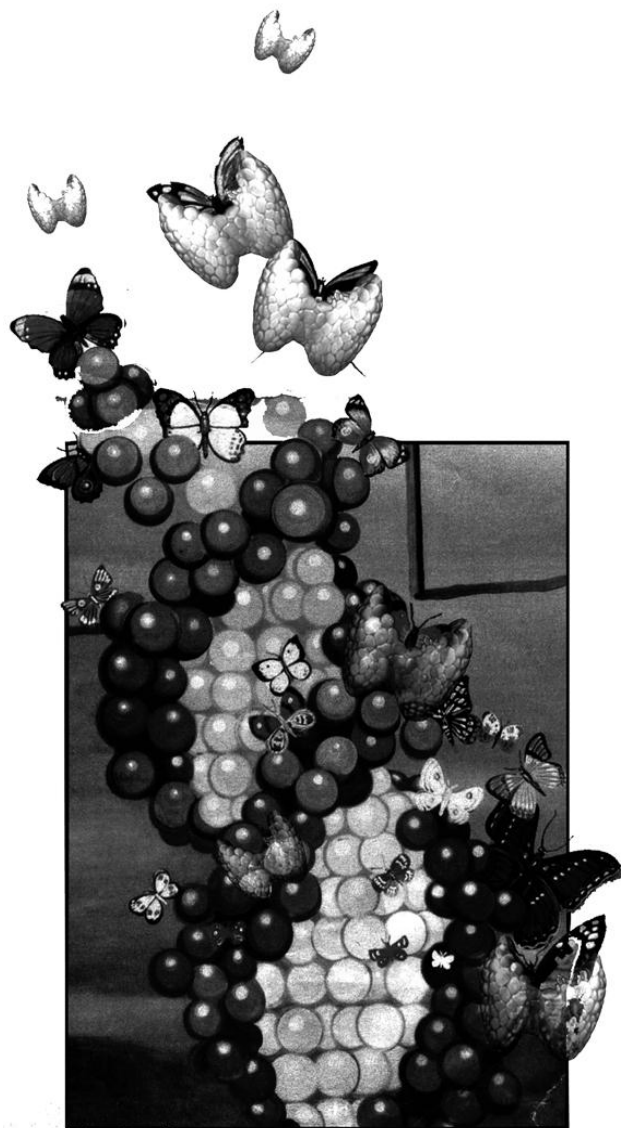
Al comienzo de esta tesis, y a diferencia de lo descrito para otras neoplasias, apenas existían publicaciones sobre estudios caso-control para la detección de GBP en carcinomas de tiroides de estirpe folicular. Además, entre las pocas existentes, tan sólo una de ellas, que identificó variantes de riesgo en el gen *TG*, incluía una serie de más de 300 pacientes (Matakidou, et al., 2004). El resto de los estudios publicados hasta esa fecha incluían un número limitado de muestras para la adecuada evaluación de las variantes estudiadas en cada caso (Cybulski, et al., 2004; Gaspar, et al., 2004; Granja, et al., 2004; Hernandez, et al., 2003; Ho, et al., 2005; Lesueur, et al., 2002; Silva, et al., 2005; Sturgis, et al., 2005), condicionando así su potencia para detectar asociaciones reales, y muy posiblemente aumentando los falsos positivos en algunas ocasiones.

Durante estos años (2006-2010), se han ido publicando un número creciente de trabajos de este tipo (revisados por Adjadj y cols., y por Landa y Robledo (Adjadj, et al., 2009; Landa and Robledo, 2011)), destacando algunos con series más amplias o mejor caracterizadas (Gudmundsson, et al., 2009; Jazdzewski, et al., 2008; Takahashi, et al., 2010; Wokolorczyk, et al., 2008). Sin embargo, en la gran mayoría de los estudios no se establecía ningún tipo de subestratificación de los pacientes en función del subtipo tumoral desarrollado, ni ensayos funcionales que determinasen el mecanismo de los genes y variantes identificadas. Estos dos aspectos, además del reclutamiento de un amplio número de muestras, han sido los pilares de nuestro trabajo.

Por todo lo anterior, nos planteamos llevar a cabo un estudio de asociación caso-control, biológicamente orientado de acuerdo al conocimiento previo del

carcinoma de tiroides de estirpe folicular. Para ello, empleamos una aproximación de genes candidatos, seleccionados de acuerdo a tres criterios: 1- genes que hubiesen mostrado una expresión diferencial entre tumores de estirpe folicular con respecto a tiroides normal (Montero-Conde, et al., 2008), 2- genes relacionados con el metabolismo del tiroides, y 3- genes localizados en puntos clave de rutas conocidas y relacionadas con la enfermedad. Con todo ello se pretendió cubrir tanto genes cuya relación con la enfermedad fuese previamente conocida, como genes cuyo papel estuviese por definir. Este tipo de aproximación, en el que se incluían a estudio genes pertenecientes a rutas relacionadas entre sí y a la enfermedad, nos permitía plantearnos un posterior estudio de epistasia, e intentar así alcanzar una visión algo más completa de los factores de riesgo genético asociados al cáncer de tiroides de célula folicular.

OBJETIVOS



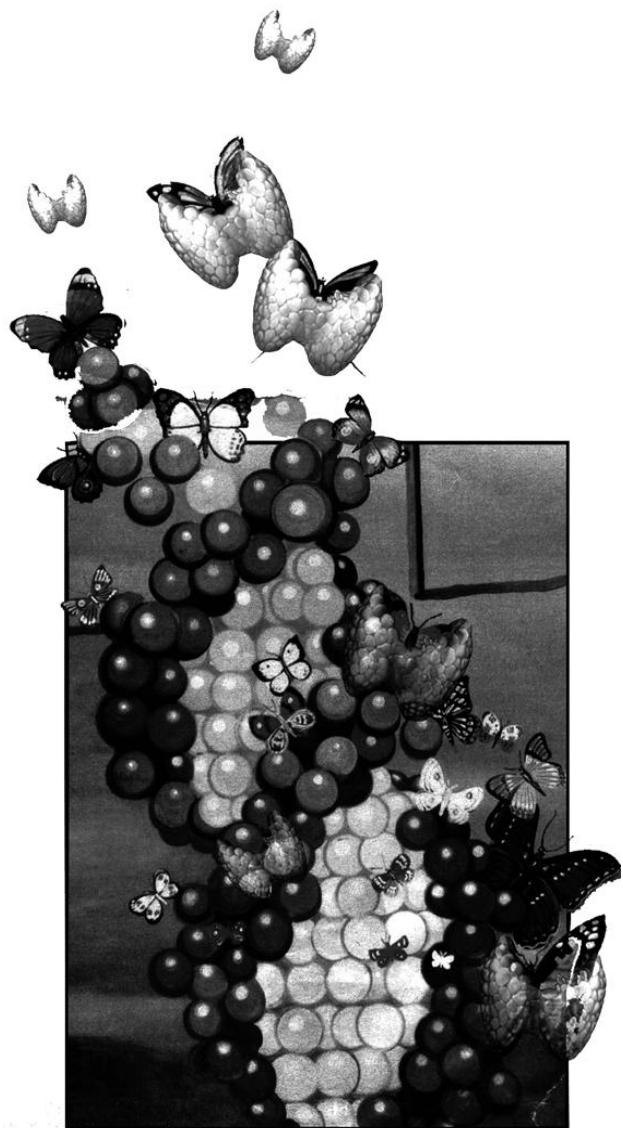
El objetivo principal de esta tesis fue la identificación de genes de baja penetrancia (GBP) relacionados con el desarrollo de carcinomas de células foliculares de tiroides. Para ello, nos planteamos los siguientes objetivos específicos:

I. Determinar si la estratificación por subtipo histológico influye o no en las señales potenciales de asociación.

II. Determinar los mecanismos subyacentes por los que las variantes identificadas y validadas actúan como factores de riesgo o protección al desarrollo de la enfermedad.

III. Identificar interacciones entre genes o epistasia, y tener así una visión más completa de la enfermedad.

MATERIAL y MÉTODOS



1. Obtención de muestras y datos clínicos

1.1. Obtención de muestras biológicas de pacientes y controles

Durante el desarrollo de este proyecto, hemos recopilado una extensa serie de pacientes con carcinomas de tiroides de estirpe folicular, fundamentalmente Carcinoma Papilar de Tiroides (CPT) y Carcinoma Folicular de Tiroides (CFT), gracias a la colaboración continua con distintos hospitales. En todos los casos, el material de trabajo fue ADN obtenido de sangre periférica o saliva, de pacientes que participaron en el estudio de forma voluntaria y rellenaron un consentimiento informado aceptando su participación en este proyecto de investigación. El establecimiento de nuevas colaboraciones en el curso del estudio nos permitió recopilar un importante número de casos y controles, que dividimos en tres series diferenciadas, cuyas características principales se detallan en los siguientes apartados y en la tabla M1.

1.1.1. Serie 1 de casos y controles

La serie 1 fue el primer grupo de muestras, obtenido entre los años 2003 y 2007, y con los que se realizó la primera etapa de nuestro estudio o fase de descubrimiento.

La serie se compuso de 615 pacientes de cáncer de tiroides derivado de célula folicular, gracias a la colaboración de 17 hospitales de la red pública del sistema sanitario, incluyendo los siguientes centros: Hospital Arnau de Vilanova (Lleida), Hospital General Universitario Gregorio Marañón (Madrid), Hospital Virgen de los Lirios de Alcoy (Alicante), Hospital Reina Sofía (Córdoba), Hospital General de Segovia, Hospital Nuestra Señora de Sonsoles (Ávila), Hospital General Universitario de Alicante, Hospital Universitario La Paz (Madrid), Hospital General Universitario Reina Sofía (Murcia), Hospital de Sabadell – Parc Taulí (Barcelona), Hospital Clínico San Carlos (Madrid), Hospital Universitario Príncipe de Asturias de Alcalá de Henares (Madrid), Hospital Universitario Sant Joan (Alicante) , Hospital de Pontevedra, Hospital Universitario Puerta de Hierro (Madrid), Hospital de Denia (Alicante) y Hospital Vega Baja de Orihuela (Alicante).

De los 615 ADNs genómicos obtenidos de pacientes, 609 mostraron la calidad y concentración suficiente para nuestro estudio. Las edades de diagnóstico estuvieron

comprendidas entre 10 y 85 años, siendo 46 años la media de edad, y la relación entre mujeres y hombres fue 4.6. De acuerdo a los datos clínicos disponibles (ver apartado 1.2), constatamos que la serie 1 de pacientes permitía tener representados los principales subtipos histológicos de cáncer de tiroides derivado de célula folicular: 520 CPT, incluyendo 305 CPT clásicos (CPTc) y 146 CPT variante folicular (CPTvf), además de 69 CFT.

Se seleccionaron 525 controles sanos, sin signos de enfermedad tiroidea, y representativos de la población española. La elección de controles se realizó de forma que variables como género y edad media estuviesen pareadas con los casos. Así, los 504 controles que fueron finalmente considerados, presentaron una edad media de 46 años y un cociente mujeres/hombres de 4.6.

1.1.2. Serie 2 de casos y controles

Durante los años 2008 y 2009, se recogió una segunda población de casos y controles procedentes de tres regiones italianas, a través de la colaboración con el Dr. Giuseppe Opocher, del *Istituto Oncologico Veneto* de Padova (Italia) y el Dr. Giuseppe Viglietto, de la *Università Magna Graecia* en Catanzaro (Italia). Esta serie 2, o serie de validación, fue empleada para validar los resultados obtenidos en población española en la primera fase del estudio, y evitar así la presencia de falsos positivos.

Se obtuvieron un total de 491 ADNs de pacientes con carcinomas de tiroides de estirpe folicular, procedentes de las regiones de Verona (n=168), Padova (n=226) y Catanzaro (n=97). De ellos, 482 fueron considerados para análisis posteriores, incluyendo 412 CPT y 44 CFT. La edad media al diagnóstico fue de 49 años, y el cociente mujer/hombre igual a 4.9.

Paralelamente, se obtuvieron 584 controles, representativos de las tres regiones geográficas de donde procedían los casos (513 del Norte de Italia – Padova y Verona; y 71 del Sur, región de Catanzaro). De ellos, 500 pasaron los estándares de calidad, presentando una edad media de 59 años y una distribución hombre/mujer de 1.6.

1.1.3. Serie 3 de casos y controles

Entre 2009 y 2010, tuvimos acceso a una tercera serie de pacientes, gracias a la colaboración con el Dr. Ricard Marcos, de la Universidad Autónoma de Barcelona. Estos casos (n=405), procedentes de cuatro hospitales de la provincia de Barcelona, se complementaron con 174 casos adicionales, recogidos durante el periodo 2008-2010 en la red de hospitales mencionados para la serie 1, pero no incluidos en ésta. En total, la serie 3 contó con 579 casos, de los cuales 538 fueron finalmente considerados, con una edad media de 42 años y un cociente mujer/hombre de 3.9. Esta serie de casos incluyó 451 CPT y 62 CFT. Se eligió un grupo de controles españoles para la serie 3, que finalmente incluyó a 540 individuos, con una edad media de 48 años y un cociente mujer/hombre igual a 3.8.

Las series 2 y 3 fueron consideradas como una sola para posteriores estudios de validación. Esta serie combinada estuvo compuesta por 1020 pacientes (862 CPT y 104 CFT), y 1040 controles (Tabla M1).

Tabla M1. Características principales de los grupos de casos y controles incluidos en el estudio

		Serie 1 (descubrimiento)	Serie 2 (validación)	Serie 3 (validación)	Serie 2+3 (validación)
Casos	N	609	482	538	1020
	Origen	España	Italia	España	Italia+España
	Edad	46	49	42	47
	Género	4.6	4.9	3.9	4.4
Controles	N	504	500	540	1040
	Origen	España	Italia	España	Italia+España
	Edad	46	59	48	53
	Género	4.6	1.6	3.8	2.4

N=número de muestras; el género se representa como el cociente entre mujeres y hombres.

1.2. Obtención de datos clínicos

De todos los pacientes se recogieron datos clínicos de forma homogénea, de acuerdo al cuestionario mostrado en la Figura M1, que incluye información sobre género y edad, antecedentes personales y familiares de cáncer, clasificación histológica (subtipo) y anatomía patológica (código TNM) del tumor, tipo de cirugía practicada, tratamiento y dosis recibida, y desarrollo de recidivas durante el seguimiento. Esta información fue posteriormente utilizada para estratificar los resultados en función del subtipo histológico tumoral, y seleccionar cuidadosamente los controles para que estuviesen pareados con los casos en la medida de lo posible.

Además, estos datos, como ya se ha comentado, nos permitieron definir los grupos de comparación según el subtipo tumoral tiroideo (CPT, CPTc, CPTvf, CFT). También, gracias a estos datos clínicos, pudimos identificar un grupo de pacientes de acuerdo a otro criterio: el comportamiento de los tumores. De esta forma, definimos un grupo de casos como "agresivos", consistente en aquéllos pacientes que presentasen metástasis en nódulos linfáticos locales ($T_xN_1M_x$) o metástasis a distancia ($T_xN_xM_1$), tanto en el diagnóstico inicial como durante el seguimiento. Este grupo de casos ascendió a 217 pacientes en la serie de descubrimiento, y a 152 para las series de validación.

Pacientes con cáncer de tiroides esporádico **Hospital:**.....

Nº Identificación paciente _____
Médico responsable _____ e-mail o tfno. contacto: _____
VARON ☐ MUJER ☐
FECHA NACIMIENTO _____
RAZA _____ SEXO _____

Antecedentes familiares

Ca. tiroideo (sí/no)
Otros tumores (especificar)

Antecedentes personales: Radioterapia cervical en la infancia (sí/no)

Cirugía

Fecha _____
Tipo de Cirugía _____

(Tiroidectomía total linfadenectomía, Tiroidectomía total, Tiroidectomía subtotal, Hemitiroidectomía, Nodulectomía)

Anatomía patológica

Tamaño del tumor _____ Estadío _____
Invasión local ☐
Metástasis ganglionar ☐
Metástasis a distancia ☐

Histología

- Carcinoma Papilar
 - puro ☐
 - folicular ☐
 - células altas ☐
 - trabecular ☐
 - Otros ☐ (especificar)
- Carcinoma Folicular
 - Mínimamente invasivo ☐
 - Invasivo ☐
- Carcinoma Hürtle ☐
- Carcinoma Medular ☐
 - Hiperplasia ☐ Único ☐ Múltiple ☐

Revisiones

- Dosis total recibida de Yodo radioactivo _____
- Recidiva sí ☐ no ☐ (Fecha) _____
 - Local ☐
 - Pulmonar ☐
 - Ósea ☐
 - Otras ☐ (especificar) _____

Figura M1. Modelo de cuestionario clínico estándar para pacientes con cáncer de tiroides esporádico

2. Extracción y procesamiento de ADN

2.1. Extracción de ADN genómico a partir de linfocitos de sangre periférica

La mayor parte de los ADNs incluidos en el estudio fueron extraídos a partir de linfocitos de sangre periférica. Para ello, la muestra de sangre se centrifugó 10 minutos a 3000 rpm, de modo que se separasen claramente las distintas fases de la sangre. Con una pipeta Pasteur estéril, se recogió cuidadosamente la capa blanquecina que contiene los linfocitos (normalmente un volumen entre 500 µl y 1 ml), y se transfirió a un tubo eppendorf, bien para su procesamiento inmediato, bien para su uso posterior, en cuyo caso se congeló a -20 °C.

Para la extracción de ADN genómico de linfocitos, se utilizaron dos métodos distintos, dependiendo de cada caso: extracción automática de ADN mediante el robot *MagNA Pure LC* (Roche, Mannheim, Alemania) y extracción manual con el kit de *FlexiGene* (QIAGEN, Valencia, CA, USA). En general, el método manual se utilizó para muestras de las que disponíamos de un volumen de sangre escaso, ya que el rendimiento final es mayor que el obtenido con los métodos automáticos.

2.1.1. Extracción automática de ADN genómico mediante *MagNA Pure LC*

La extracción mediante este sistema se realiza con un robot, capaz de procesar simultáneamente 32 muestras, y consiste en los siguientes pasos (calculados por muestra):

1. Partiendo de un volumen de 1 ml de muestra (500 µl de linfocitos + 500 µl de NaCl 0.9%), el sistema añade 115 µl de proteinasa K para la digestión de proteínas.
2. Añadir 838 µl de tampón de lisis/unión para favorecer la lisis celular y generar las condiciones salinas apropiadas para la posterior retención del ADN.
3. Añadir 544 µl del reactivo MGP, que retiene el ADN a través de las partículas cristalinas magnéticas que contiene.
4. Añadir 2.42 ml de tampón de lavado I para eliminar sustancias no unidas a las partículas magnéticas, tales como proteínas, membranas celulares e inhibidores de amplificación (heparina).
5. Añadir 1.28 ml de tampón de lavado II para retirar restos celulares y reducir la concentración salina.
6. Añadir 831 µl de tampón de lavado III para eliminar restos de etanol.
7. Resuspender el ADN purificado en 200 µl del tampón de elución, que se calienta previamente para favorecer el proceso.

2.1.2. Extracción de ADN genómico con el kit de *FlexiGene*

Para extraer el ADN con el kit de Flexigene (QIAGEN), se realizan los siguientes pasos, que están referidos a un volumen inicial de linfocitos ("buffy coat") de 500 µl:

1. Si se parte de una capa de linfocitos congelada, descongelarla rápidamente en un baño a 37 °C con agitación suave.
2. Añadir 1250 µl de tampón FG1 y mezclar invirtiendo el tubo cinco veces.
3. Centrifugar 20 s a 10000 *g*.
4. Descartar el sobrenadante y dejar el tubo abierto e invertido sobre una porción de papel secante, teniendo cuidado de no perder el pellet.
5. Añadir 500 µl de tampón FG2 y 5 µl de la proteasa suministrada en el kit.
6. Cerrar el tubo y agitar de inmediato en un vortex hasta que el pellet esté completamente homogeneizado. En el caso de que se observen trazas de pellet con aspecto gelatinoso, añadir otros 30 µl de tampón FG2 y agitar de nuevo.
7. Asegurarse de que el líquido está en el fondo del tubo (con una breve centrifugación si es preciso), e incubarlo a 65 °C durante 5 min en un bloque térmico (Thermomixer comfort, Eppendorf). La solución adquiere un color verde oscuro.
8. Añadir 500 µl de isopropanol absoluto y mezclar bien por inversión, hasta que el ADN precipitado se vuelva visible en forma de hebras blanquecinas o formando un agregado con aspecto de medusa.
9. Centrifugar 3 minutos a 10000 *g*.
10. Descartar el sobrenadante y eliminar los restos de líquido con papel absorbente. El pellet de ADN se verá como un pequeño depósito blanco en el fondo del tubo.
11. Lavar el pellet con 500 µl de etanol al 70%, agitar con un vortex y volver a centrifugar a 10000 *g* durante 3 min.
12. Eliminar el sobrenadante y secar el pellet dejando el tubo abierto durante 10 minutos.
13. Resuspender el ADN en 200 µl de tampón FG3, agitar brevemente en un vortex e incubarlo a 65 °C durante al menos 30 min en un bloque térmico.

2.2. Extracción de ADN genómico a partir de saliva

Durante el desarrollo de este proyecto, un nuevo kit comercial, el *Oragene DNA Self-Collection Kit* (DNA Genotek, Ottawa, Canada), apareció en el mercado, permitiendo la extracción de ADN genómico a partir de las células del epitelio bucal contenidas en la saliva. Dada la facilidad en la recogida de la muestra, su estabilidad a temperatura ambiente durante largo tiempo y la calidad del ADN resultante, medida tanto en términos de concentración como de integridad, fuimos sustituyendo progresivamente la recogida de muestras de sangre de pacientes por muestras de saliva. El protocolo de extracción de ADN a partir de saliva se detalla a continuación:

1. Agitar los kits para que la saliva se mezcle con la solución estabilizadora, e incubar la mezcla en una estufa a 50 °C durante al menos 2 h, aunque también puede incubarse durante toda la noche si es más conveniente.
2. Transferir el contenido a un tubo Falcon de 15 ml y añadir un volumen 1:25 del reactivo *Oragene Purifier* (unos 160 µl de reactivo para 4 ml de saliva)
3. Mezclar bien por inversión e incubar en hielo durante 10 min. La muestra se vuelve turbia porque las impurezas precipitan.
4. Centrifugar 20 min a 4000 *g* a 4 °C.
5. Pipetear con cuidado el sobrenadante limpio de cada tubo y recogerlo en otro Falcon.
6. Añadir 4 ml (igual volumen) de etanol absoluto y mezclar por inversión. El ADN se verá como hebras blanquecinas o formando una medusa.
7. Incubar 10 min a temperatura ambiente para que el ADN precipite, y a continuación, centrifugar 10 min a 1100 *g* (3500 rpm).
8. Descartar el sobrenadante, tratando de evaporar las últimas trazas de etanol dejando el tubo abierto durante unos minutos.
9. Resuspender el pellet de ADN en 500 µl de agua. Favorecer la disolución total del ADN, incubando durante 2 h en un baño a 50 °C y/o durante toda la noche a temperatura ambiente, preferiblemente con agitación.

2.3. Amplificación del genoma completo mediante GenomiPhi™

En algunas muestras de las series 2 y 3, de los que la muestra de ADN era escasa, se llevó a cabo una amplificación del genoma completo mediante el sistema *GenomiPhi™ V2 DNA Amplification Kit* (GE Healthcare). El protocolo consta de los siguientes pasos:

1. Desnaturalizar el ADN problema, mezclando, en un tubo de 0.2 ml, 1 µl del mismo con 9 µl de *sample buffer*, y calentándolo a 95 °C durante 3 min. Enfriar en hielo.
2. Para cada muestra a amplificar, preparar una mezcla de 9 µl de *reaction buffer* y 1 µl del *enzyme mix*. Añadirlo a cada muestra fría.
3. Incubar la mezcla a 30 °C durante 2 h para que la amplificación tenga lugar.
4. Inactivar la enzima calentando la mezcla a 65 °C durante 10 min.

Los productos de amplificación de todo el genoma requieren una purificación posterior, que consiste en los siguientes pasos:

1. Preparar un tampón con volúmenes iguales (1:1) de acetato sódico y EDTA, a una concentración final de 1.5 M y 250 mM, respectivamente.
2. Transferir el amplificado a un eppendorf de 1.5 ml, y añadir, por muestra, 20 µl de agua Braun y 4 µl del tampón acetato sódico:EDTA
3. Añadir a cada muestra 100 µl de etanol absoluto, mezclar con un vortex durante 1 min, y centrifugar a 13000 rpm durante 30 min.
4. Descartar el sobrenadante y lavar el pellet de ADN con 100 µl de etanol al 70 %. Centrifugar a 13000 rpm durante 30 min.
5. Eliminar todo el etanol, dejando evaporar la muestra toda la noche.
6. Resuspender el ADN en 50 µl de agua Braun, agitando la muestra en un termobloque a 37 °C durante 2 h, aproximadamente.

La cantidad de ADN obtenida mediante GenomiPhi™ se midió, al igual que el resto de muestras, mediante el reactivo PicoGreen (apartado 2.4), obteniéndose una concentración media de 50 ng/µl, aproximadamente. Los ADNs amplificados por GenomiPhi™ se emplearon para el genotipado mediante las plataformas TaqMan® y KASPar (apartados 4.2 y 4.3), que pueden trabajar con ADNs de partida amplificados por este sistema. Antes de considerar que un ADN amplificado era representativo del correspondiente ADN original, se genotiparon en el mismo ensayo algunos polimorfismos para ambos ADNs. Sólo en el caso de obtener idénticos resultados, se utilizó el ADN amplificado en las fases de validación.

2.4. Cuantificación del ADN genómico mediante PicoGreen

La concentración de todos los ADNs incluidos en el estudio se cuantificó empleando el sistema *Quant-iT PicoGreen dsDNA Reagent* (Invitrogen, Eugene, OR, USA). Este reactivo es un agente intercalante, capaz de unirse al ADN y emitir fluorescencia de una forma directamente proporcional a la cantidad de ADN presente en la reacción. El método se basa en la comparación del valor de fluorescencia de la muestra problema con los valores obtenidos por muestras de concentración creciente y conocida.

Brevemente, el protocolo consiste en la preparación de una recta patrón con diluciones seriadas de ADN genómico (Clontech, Mountain View, CA, USA), cubriendo un rango de concentraciones entre 0 y 160 ng/μl. Estos ADNs se dispensan en duplicado en una placa opaca de 96 pocillos (Falcon, BD Biosciences, San Jose, CA, USA), y junto a ellos, se pipetea 2 μl de cada una de las muestras problema. El reactivo PicoGreen, convenientemente protegido de la luz, se prepara a una concentración 1:400 en buffer TE 1X (10 mM Tris, 1 mM EDTA, pH 7.5), y se dispensan 195 μl de esta dilución a cada pocillo. La placa se agita brevemente en un agitador de placas, se incuba en oscuridad durante cinco minutos y se centrifuga durante un minuto. A continuación, se mide la fluorescencia de cada muestra a una longitud de onda de 520 nm tras excitación a 480 nm, empleando el aparato *DTX 800 Multimode Detector* (Beckman Coulter, Fullerton, CA, USA). Con los datos de fluorescencia, se construye una recta patrón y, tras comprobar que existe una buena correlación lineal de los datos ($r^2 \rightarrow 1$), se interpolan las muestras problema para calcular su concentración, como se muestra en la Figura M2.

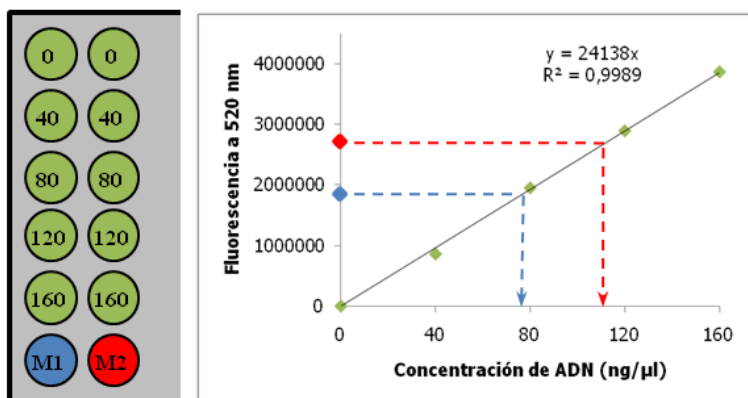


Figura M2. Cuantificación de ADN genómico mediante PicoGreen. La recta patrón se construye representando gráficamente fluorescencia *versus* concentración de duplicados de ADN de concentración conocida (en verde). La concentración de las muestras problema (M1 y M2, en azul y rojo) se calcula por interpolación.

3. Selección de genes y polimorfismos para el estudio de asociación

3.1 Selección de genes

Se eligió una estrategia de genes candidatos, donde la selección tuvo en cuenta el sentido biológico de los mismos, y de acuerdo a los siguientes criterios:

1. En primer lugar, se seleccionaron genes que hubiesen presentado expresión transcripcional diferencial y estadísticamente significativa al comparar distintos subtipos de carcinomas tiroideos de estirpe folicular y tejidos normales. Estos resultados se habían obtenido previamente en nuestro laboratorio (Montero-Conde, et al., 2008). Estos datos fueron complementados con aquéllos disponibles en bases de datos públicas. En concreto, empleamos la información del *SAGE Anatomic Viewer* (<http://cgap.nci.nih.gov/SAGE/AnatomicViewer>), que permite conocer la expresión diferencial de muchos genes en tejido tumoral (tiroideo, en nuestro caso) y su contrapartida normal.
2. Genes relacionados con el metabolismo tiroideo, concretamente los relacionados con las células foliculares de tiroides, tratando de seleccionar genes esenciales para la función de estas células, fundamentalmente dedicadas a la producción de hormonas tiroideas, con la idea de que alteraciones en esos actores podrían jugar un papel en la predisposición a desarrollar carcinomas de tiroides.
3. Por último, se eligieron rutas previamente conocidas como esenciales para el funcionamiento del metabolismo, seleccionando genes con un papel clave en ellas, bien por estar situados en el inicio de la misma (receptores, transductores de señales, etc.), bien por localizarse en una posición importante, por ejemplo, interactuando con varias proteínas de la vía. Para la selección de genes en rutas, además de información propia, empleamos tres recursos bioinformáticos especializados: la colección de rutas metabólicas *Biocarta* (<http://www.biocarta.com/genes/index.asp>), la colección de rutas de la Enciclopedia de Genes y Genomas de Kyoto, *KEGG Pathways* (<http://www.genome.jp/kegg/pathway.html>) y el programa *Pathway Studio 4*, versión de evaluación (Ariadne Genomics, MD, USA).

Con todos estos criterios, elaboramos una lista, que inicialmente redujimos a 219 genes candidatos. En principio, se seleccionaron tanto genes que codifican proteínas, como aquéllos que codifican microRNAs, gracias a la herramienta miRbase (<http://microrna.sanger.ac.uk/sequences/>), pero estos últimos fueron finalmente descartados por no encontrarse variantes polimórficas en ellos, al menos en el momento en el que este estudio fue planteado. Se estableció entonces una lista de genes codificantes, que fueron ordenados de acuerdo a la importancia biológica, y finalmente se seleccionaron un total de 97 genes para el genotipado. Los genes elegidos y los criterios de selección para cada uno de ellos están resumidos en la tabla M2.

Además, paralelamente a este proyecto, se seleccionó el gen *CDKN1B* (p27), dado su relación demostrada con las neoplasias de tiroides de célula folicular (Chiappetta, et al., 2007; Chu, et al., 2008; Khoo, et al., 2002; Viglietto, et al., 2002), para estudiar sus variantes polimórficas en relación con este cáncer.

Tabla M2. Genes seleccionados para el estudio y polimorfismos seleccionados en cada uno de ellos

Genes seleccionados para el estudio					Criterios de selección de genes			Polimorfismos genotipados				
#	Gen (HUGO)	Cr	ID (NCBI)	Nombre oficial (NCBI Gene)	Criterio 1: expresión diferencial	Criterio 2: metabolismo tiroideo	Criterio 3: ruta metabólica	tag	no sin	ESE	TFBS	Tot
1	<i>DIO1</i>	1	1733	Deiodinase, iodothyronine, type I	FDR<0.15 en CPT+CFT	Acción de hormonas tiroideas	#	2	0	1	0	3
2	<i>PTH1H</i>	12	5744	Parathyroid hormone-like hormone	FDR<0.15 en CPT+CFT	Paratiroides	#	6	0	1	0	7
3	<i>PTH1R</i>	3	5745	Parathyroid hormone 1 receptor	FDR<0.15 en CPT+CFT	Paratiroides	#	8	0	0	1	9
4	<i>THADA</i>	2	63892	Thyroid adenoma associated	FDR<0.15 en CPT+CFT	Reordenamiento con PPARG	#	25	3	1	1	30
5	<i>THRA</i>	17	7067	Thyroid hormone receptor, alpha (v-erb-a oncogene homolog, avian)	FDR<0.15 en CPT+CFT	Recepción de hormonas tiroideas	#	1	0	0	0	1
6	<i>MED13</i>	17	9969	Mediator complex subunit 13 (=THRAP1)	ns (FDR>0.15)	Recepción de hormonas tiroideas	#	5	0	0	0	5
7	<i>TRIP11</i>	14	9321	Thyroid hormone receptor interactor 11	FDR<0.15 en CPT+CFT	Recepción de hormonas tiroideas	#	13	2	0	0	15
8	<i>TRIP12</i>	2	9320	Thyroid hormone receptor interactor 12	ns (FDR>0.15)	Recepción de hormonas tiroideas	#	8	0	2	1	11
9	<i>CCL5</i>	17	6352	Chemokine (C-C motif) ligand 5	ns (FDR>0.15)	Autoinmunidad en tiroides	#	2	0	1	0	3
10	<i>PRKAR1A</i>	17	5573	Protein kinase, cAMP-dependent, regulatory, type I, alpha	ns (FDR>0.15)	Reordenamiento RET/PTC2 en CPT	#	3	0	4	1	8
11	<i>PAX8</i>	2	7849	Paired box gene 8	ns (FDR>0.15)	Reordenamiento con PPARG en CFT	#	19	0	2	0	21
12	<i>TPO</i>	2	7173	Thyroid peroxidase	FDR<0.15 en CPT+CFT	Síntesis de hormonas tiroideas	Síntesis de horm. tiroideas	9	3	1	2	15
13	<i>KCNIP3</i>	2	30818	Kv channel interacting protein 3, calsenilin (=DREAM)	Sin datos	Represor de FOXE1	Ruta TSH-cAMP en tiroides	10	0	0	0	10
14	<i>NCOA4</i>	10	8031	Nuclear receptor coactivator 4	Sin datos	Reordenamiento RET/PTC3 en CPT	#	9	0	0	0	9
15	<i>CCDC6</i>	10	8030	Coiled-coil domain containing 6	Sin datos	Reordenamiento RET/PTC1 en CPT	#	45	1	3	0	49
16	<i>NTRK1</i>	1	4914	Neurotrophic tyrosine kinase, receptor, type 1	Sin datos	Reordenamientos en CPT	Ruta MAPK; Apoptosis	11	2	3	1	17

#	Gen (HUGO)	Cr	ID (NCBI)	Nombre oficial (NCBI Gene)	Criterio 1: expresión diferencial	Criterio 2: metabolismo tiroideo	Criterio 3: ruta metabólica	tag	no sin	ESE	TFBS	Tot
17	<i>SLC26A4</i>	7	5172	Solute carrier family 26, member 4	Sin datos	Síntesis de hormonas tiroideas	Síntesis de horm. tiroideas	15	1	0	0	16
18	<i>SLC5A8</i>	12	160728	Solute carrier family 5 (iodide transporter), member 8	Sin datos	Síntesis de hormonas tiroideas	Síntesis de horm. tiroideas	14	2	1	1	18
19	<i>TG</i>	8	7038	Thyroglobulin	Sin datos	Síntesis de hormonas tiroideas	Síntesis de horm. tiroideas	28	7	5	2	42
20	<i>DUOX1</i>	15	53905	Dual oxidase 1	Sin datos	Síntesis de hormonas tiroideas	Síntesis de horm. tiroideas	7	2	4	3	16
21	<i>TSHB</i>	1	7252	Thyroid stimulating hormone, beta	Sin datos	Subunidad específica de la TSH	Ruta TSH-cAMP en tiroides	7	1	0	4	12
22	<i>TSHR</i>	14	7253	Thyroid stimulating hormone receptor	Sin datos	Receptor de la TSH	Ruta TSH-cAMP en tiroides	13	2	2	1	18
23	<i>NKX2-1</i>	14	7080	NK2 homeobox 1 (thyroid transcription factor 1)	Sin datos	Factor de transcripción tiroideo	Ruta TSH-cAMP en tiroides	2	0	0	0	2
24	<i>FOXE1</i>	9	2304	Forkhead box E1 (thyroid transcription factor 2)	ns (FDR>0.15)	Factor de transcripción tiroideo	Ruta TSH-cAMP en tiroides	4	0	1	2	7
25	<i>CGA</i>	6	1081	Glycoprotein hormones, alpha polypeptide	Sin datos	Subunidad común de la TSH	Ruta TSH-cAMP en tiroides	7	0	2	4	13
26	<i>RTN3</i>	11	10313	Reticulon 3	FDR<0.15 en CPT	#	#	2	2	2	0	6
27	<i>CTHRC1</i>	8	115908	Collagen triple helix repeat containing 1	FDR<0.15 en CPT	#	#	5	1	2	2	10
28	<i>FGFR4</i>	5	2264	Fibroblast growth factor receptor 4	FDR<0.15 en CPT	#	Ruta MAPK	3	2	1	0	6
29	<i>LGALS1</i>	22	3956	Lectin, galactoside-binding, soluble, 1 (galectin 1)	FDR<0.15 en CPT	#	#	2	0	0	1	3
30	<i>S100A10</i>	1	6281	S100 calcium binding protein A10	FDR<0.15 en CPT	#	Metástasis	6	0	0	1	7
31	<i>STK17B</i>	2	9262	Serine/threonine kinase 17b (apoptosis-inducing)	FDR<0.15 en CPT	#	Apoptosis	5	0	0	0	5
32	<i>TCEB3</i>	1	6924	Transcription elongation factor B (SIII), polypeptide 3 (elongin A)	FDR<0.15 en CPT	#	#	4	1	0	0	5
33	<i>SIAH2</i>	3	6478	Seven in absentia homolog 2 (Drosophila)	FDR<0.15 en CPT	#	#	0	0	1	1	2
34	<i>CSF2</i>	5	1437	Colony stimulating factor 2 (granulocyte-macrophage)	FDR<0.15 en CPT+CFT	#	Ruta Ca ²⁺ /PKC	6	1	0	2	9

#	Gen (HUGO)	Cr	ID (NCBI)	Nombre oficial (NCBI Gene)	Criterio 1: expresión diferencial	Criterio 2: metabolismo tiroideo	Criterio 3: ruta metabólica	tag	no sin	ESE	TFBS	Tot
35	STAT2	12	6773	Signal transducer and activator of transcription 2, 113kDa	FDR<0.15 en CPT+CFT	#	Ruta JAK/STAT	1	1	0	0	2
36	JAK1	1	3716	Janus kinase 1 (a protein tyrosine kinase)	FDR<0.15 en CPT+CFT	#	Ruta JAK/STAT	14	0	4	1	19
37	GHR	5	2690	Growth hormone receptor	FDR<0.15 en CPT+CFT	#	Ruta AKT/PI3K; JAK/STAT	6	1	2	0	9
38	SMO	7	6608	Smoothened homolog (Drosophila)	FDR<0.15 en CPT+CFT	#	Ruta Sonic-Hedgehog	1	0	6	0	7
39	SMAD7	18	4092	SMAD, mothers against DPP homolog 7 (Drosophila)	FDR<0.15 en CPT+CFT	#	Ruta TGFB-Smad	12	0	0	2	14
40	IRS1	2	3667	Insulin receptor substrate 1	FDR<0.15 en CPT+CFT	#	Insulina y fact crecimiento	0	1	2	0	3
41	MYC	8	4609	V-myc myelocytomatosis viral oncogene homolog (avian)	FDR<0.15 en CPT+CFT	#	Ruta Ca ²⁺ /PKC	0	1	1	0	2
42	CXCR4	2	7852	Chemokine (C-X-C motif) receptor 4	Sí, otros grupos	#	Ruta MAPK	5	0	1	0	6
43	KIT	4	3815	V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	Sí, otros grupos	Expresión alterada en CPT	#	26	1	1	0	28
44	JAK2	9	3717	Janus kinase 2 (a protein tyrosine kinase)	ns (FDR>0.15)	#	Ruta JAK/STAT	12	1	2	2	17
45	STAT3	17	6774	Signal transducer and activator of transcription 3	ns (FDR>0.15)	#	Ruta JAK/STAT	6	0	3	2	11
46	SRC	20	6714	V-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)	Sin datos	#	Ruta JAK/STAT	8	0	1	1	10
47	BAD	11	572	BCL2-antagonist of cell death	FWER<0.05	#	Ruta AKT/PI3K, Apoptosis	2	0	0	1	3
48	PPARG	3	5468	Peroxisome proliferative activated receptor, gamma	ns (FDR>0.15)	Reordenamiento PAX8 en CFT	Ruta de PPAR	17	1	1	0	19
49	SMAD4	18	4089	SMAD, mothers against DPP homolog 4 (Drosophila)	ns (FDR>0.15)	#	Ruta TGFB-Smad	3	0	0	0	3
50	RAF1	3	5894	V-raf-1 murine leukemia viral oncogene homolog 1	Sin datos	#	Ruta AKT/PI3K; MAPK	10	0	2	1	13
51	IL6ST	5	3572	Interleukin 6 signal transducer (gp130, oncostatin M receptor)	ns (FDR>0.15)	#	Respuesta inmune	6	1	0	1	8
52	CSGALNACT1	8	55790	Chondroitin beta1,4 N-acetylgalactosaminyltransferase	FWER<0.05	#	#	11	2	2	2	17
53	MAPK1	22	5594	Mitogen-activated protein kinase 1	Sin datos	#	Ruta MAPK	8	0	0	1	9

#	Gen (HUGO)	Cr	ID (NCBI)	Nombre oficial (NCBI Gene)	Criterio 1: expresión diferencial	Criterio 2: metabolismo tiroideo	Criterio 3: ruta metabólica	tag	no sin	ESE	TFBS	Tot
54	SHH	7	6469	Sonic hedgehog homolog (Drosophila)	Sin datos	#	Ruta Sonic-Hedgehog	3	0	1	1	5
55	PTCH1	9	5727	Patched 1 (Patched homolog (Drosophila))	Sin datos	#	Ruta Sonic-Hedgehog	0	1	2	0	3
56	ARRB2	17	409	Arrestin, beta 2	Sin datos	#	Ruta MAPK	1	0	0	1	2
57	MAP3K12	12	7786	Mitogen-activated protein kinase kinase 12	Sin datos	#	Ruta MAPK	5	0	0	0	5
58	STK17A	7	9263	Serine/threonine kinase 17a (apoptosis-inducing)	Sin datos	#	Apoptosis	13	2	2	0	17
59	RHOG	11	391	Ras homolog gene family, member G (rho G)	Sin datos	#	Ruta MAPK	12	0	1	0	13
60	ITGB2	21	3689	Integrin, beta 2	FDR<0.15 en CPT+CFT	#	Adhesión; Citoesqueleto	17	0	1	0	18
61	ITPR1	3	3708	Inositol 1,4,5-triphosphate receptor, type 1	FWER<0.05	#	Ruta Ca ²⁺ /PKC	0	0	4	0	4
62	FIGF	X	2277	C-fos induced growth factor (=VEGFD)	FDR<0.15 en CPT+CFT	#	Hipoxia y angiogénesis	9	0	1	1	11
63	FGFR1	8	2260	Fibroblast growth factor receptor 1	Sin datos	#	Ruta MAPK	9	0	3	0	12
64	ACHE	7	43	Acetylcholinesterase (Yt blood group)	FDR<0.15 en CPT+CFT	#	Génesis de β-amiloide	0	2	0	0	2
65	AMH	19	268	Anti-Mullerian hormone	Sin datos	#	Ruta TGFB-Smad	0	1	0	0	1
66	ARNT	1	405	Aryl hydrocarbon receptor nuclear translocator	Sin datos	#	Hipoxia y angiogénesis	0	1	0	0	1
67	CDYL	6	9425	Chromodomain protein, Y-like	FWER<0.05	#	#	0	1	0	0	1
68	EP300	22	2033	E1A binding protein p300	Sin datos	#	Ruta TGFB-Smad	0	2	0	0	2
69	ERO1LB	1	56605	ERO1-like beta (S. cerevisiae)	FWER<0.05	#	#	0	1	0	0	1
70	FN1	2	2335	Fibronectin 1	FDR<0.15 en CPT+CFT	#	Ruta AKT/PI3K; Citoesqueleto	0	1	0	0	1
71	IFNAR1	21	3454	Interferon (alpha, beta and omega) receptor 1	FDR<0.15 en CPT+CFT	#	Ruta JAK/STAT	0	1	0	0	1
72	IGFBP1	7	3484	Insulin-like growth factor binding protein 1	FDR<0.15 en CPT+CFT	#	Insulina y fact de crecimiento	0	1	0	0	1

#	Gen (HUGO)	Cr	ID (NCBI)	Nombre oficial (NCBI Gene)	Criterio 1: expresión diferencial	Criterio 2: metabolismo tiroideo	Criterio 3: ruta metabólica	tag	no sin	ESE	TFBS	Tot
73	<i>IKBKB</i>	8	3551	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase β	FDR<0.15 en CPT+CFT	#	Ruta AKT/PI3K; Ruta NF κ B	0	2	0	0	2
74	<i>IL3</i>	5	3562	Interleukin 3 (colony-stimulating factor, multiple)	Sin datos	#	Apoptosis	0	1	0	0	1
75	<i>IL31RA</i>	5	133396	Interleukin 31 receptor A	Sin datos	#	Ruta JAK/STAT	0	1	0	0	1
76	<i>IL3RA</i>	X	3563	Interleukin 3 receptor, alpha (low affinity)	Sin datos	#	Apoptosis	0	2	0	0	2
77	<i>ITGA1</i>	5	3672	Integrin alpha-1 (Laminin and collagen receptor) (VLA-1)	Sin datos	#	Adhesión	0	3	0	0	3
78	<i>ITGA3</i>	17	3675	Integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3 receptor)	FDR<0.15 en CPT+CFT	#	Adhesión	0	2	0	0	2
79	<i>ITPR3</i>	6	3710	Inositol 1,4,5-triphosphate receptor, type 3	Sin datos	#	Ruta Ca ²⁺ /PKC	0	1	0	0	1
80	<i>GPR98</i>	5	84059	G protein-coupled receptor 98 (=MASS1)	FWER<0.05	#	#	0	12	0	0	12
81	<i>MATN2</i>	8	4147	Matrilin 2	FWER<0.05	#	#	0	3	0	0	3
82	<i>MLH1</i>	3	4292	MutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)	FDR<0.15 en CPT+CFT	#	Reparación del ADN	0	1	0	0	1
83	<i>MMP7</i>	11	4316	Matrix metalloproteinase 7 (matrilysin, uterine)	FDR<0.15 en CPT+CFT	#	Ruta AKT/PI3K	0	2	0	0	2
84	<i>MSH3</i>	5	4437	MutS homolog 3 (E. coli)	FDR<0.15 en CPT+CFT	#	Reparación del ADN	0	2	0	0	2
85	<i>NCOR2</i>	12	9612	Nuclear receptor co-repressor 2	FWER<0.05	#	Ruta Notch	0	2	0	0	2
86	<i>PLA2R1</i>	2	22925	Phospholipase A2 receptor 1, 180kDa	FWER<0.05	#	#	0	2	0	0	2
87	<i>PLAT</i>	8	5327	Plasminogen activator, tissue	ns (FDR>0.15)	#	Génesis de β -amiloide	0	2	0	0	2
88	<i>PLAU</i>	10	5328	Plasminogen activator, urokinase	FDR<0.15 en CPT+CFT	#	Génesis de β -amiloide	0	2	0	0	2
89	<i>PMS2</i>	7	5395	PMS2 postmeiotic segregation increased 2 (S. cerevisiae)	FDR<0.15 en CPT+CFT	#	Reparación del ADN	0	2	0	0	2
90	<i>SNX19</i>	11	399979	Sorting nexin 19	FWER<0.05	#	#	0	3	0	0	3
91	<i>HNF1A</i>	12	6927	HNF1 homeobox A (=TCF1)	Sin datos	#	Ruta de Wnt y GHR	0	3	0	0	3

#	Gen (HUGO)	Cr	ID (NCBI)	Nombre oficial (NCBI Gene)	Criterio 1: expresión diferencial	Criterio 2: metabolismo tiroideo	Criterio 3: ruta metabólica	tag	no sin	ESE	TFBS	Tot
92	<i>TNFRSF11B</i>	8	4982	Tumor necrosis factor receptor superfamily, member 11b	FWER<0.05	#	Factor de necrosis tumoral	0	1	0	0	1
93	<i>TYK2</i>	19	7297	tyrosine kinase 2	ns (FDR>0.15)	#	Ruta JAK/STAT	0	1	0	0	1
94	<i>UCP1</i>	4	7350	Uncoupling protein 1 (mitochondrial, proton carrier)	Sin datos	#	Ruta cAMP/PKA	0	1	0	0	1
95	<i>VTN</i>	17	7448	Vitronectin	Sin datos	#	Moléculas de adhesión	0	1	0	0	1
96	<i>VWF</i>	12	7450	Von Willebrand factor	FDR<0.15 en CPT+CFT	#	Adhesión	0	4	0	0	4
97	<i>WT1</i>	11	7490	Wilms tumor 1	Sin datos	#	Insulina y fact de crecimiento	0	1	0	0	1
TOTALES:								523	114	83	48	768
#	<i>CDKN1B</i>	12	1027	Cyclin-dependent kinase inhibitor 1B (p27, Kip1)	Sí, otros grupos	Relacionado con CPT	Ciclo celular; Ruta AKT/PI3K	0	1	0	2	3

Los genes están ordenados según su orden de selección; FDR (*False Discovery Rate*) y FWER (*Family Wise Error Rate*) son parámetros de corrección múltiple empleados en arrays de expresión; Abreviaturas: tag= tagSNP; no sin= SNP no sinónimo; ESE= SNP localizado en *Exonic Splicing Enhancers*; TFBS= SNP en *Transcription Factor Binding Sites*; Tot: número total de SNPs por gen; ns= no significativo por expresión.

3.2 Selección de SNPs.

La lista final de 97 genes candidatos se representaron mediante un total de 768 polimorfismos tipo SNP. Estos SNPs fueron, bien intragénicos, bien localizados en región promotora, hasta 10 kilobases (kb) antes del inicio de la transcripción, o bien localizados en la región *downstream*, hasta 2 kb después del final de la transcripción. Los códigos, posiciones y frecuencias de los polimorfismos fueron obtenidos de las bases de datos públicas *NCBI SNP database*, *Build 126* (<http://www.ncbi.nlm.nih.gov/snp>) y *Ensembl*, versión 38, abril 2006 (<http://www.ensembl.org/index.html>).

Para la selección de SNPs, se aplicaron distintos criterios de una forma secuencial, como se detalla en los siguientes apartados.

3.2.1. Criterios de selección de SNPs relativos a su localización

La estrategia de selección de polimorfismos fue doble. Por una parte, empleamos una aproximación indirecta, utilizando los SNPs extraídos del Proyecto HapMap como marcadores, o tagSNPs. Por otra parte, utilizamos una estrategia más directa, seleccionando variantes con un posible papel funcional.

3.2.1.1. SNPs utilizados como marcadores o tagSNPs

El Proyecto HapMap (Consortium, 2003) (<http://hapmap.ncbi.nlm.nih.gov/>), fue un estudio internacional que permitió identificar polimorfismos en todo el genoma y la forma en la que se heredan. De esta manera, se construyeron mapas genómicos de desequilibrio de ligamiento (también llamado LD, del inglés *Linkage Disequilibrium*) o, lo que es lo mismo, mapas de regiones que se heredan juntas, debido a su baja frecuencia de recombinación. Esta información es muy útil en los estudios de asociación, porque permite seleccionar los llamados tagSNPs, que son polimorfismos que representan o "capturan" toda la variación genética de una región LD.

Para nuestro estudio, empleamos la información de HapMap (versión 21a, phase II, <http://hapmap.ncbi.nlm.nih.gov/>) referida a la población de origen caucásico (HapMap-CEU), y seleccionamos, para cada gen, aquellos tagSNPs que representasen

adecuadamente la variación contenida en ellos, en función de su patrón específico de LD. Para ello, empleamos el programa *Haploview* (versión 3.31, junio de 2006 (<http://www.broad.mit.edu/mpg/haploview>)), que permite manejar la información disponible en HapMap, produciendo representaciones similares a la mostrada en la figura M3. Los tagSNP elegidos nos permitieron, además, realizar un análisis de estas regiones para identificar los haplotipos de riesgo.

3.2.1.2. SNPs potencialmente funcionales

Mediante el empleo de colecciones de herramientas bioinformáticas de predicción de funcionalidad, principalmente *PupaSuite* (Conde, et al., 2006) (<http://pupasuite.bioinfo.cipf.es/>), y posteriormente *F-SNP* (Lee and Shatkay, 2008) (<http://compbio.cs.queensu.ca/F-SNP/>), se seleccionaron SNPs con alguno de los siguientes tipos de efecto:

1. Codificantes no sinónimos. Polimorfismos, localizados en exones, que producen un cambio de aminoácido en la proteína resultante.
2. SNPs en sitios de unión de factores de transcripción (TFBS – *Transcription Factor Binding Sites*). SNPs situados en regiones del promotor, y a los que las herramientas de predicción *in silico* les asignan un papel como dianas de unión de factores que promueven o inhiben la transcripción del gen correspondiente.
3. SNPs en sitios de reguladores del editado o *splicing* del ARN (ESE – *Exonic Splicing Enhancers*). Variantes localizadas en secuencias de unión de proteínas que intervienen potenciando el procesado del ARN mensajero (ARNm).

La figura M3 muestra esquemáticamente la localización típica de los distintos tipos de polimorfismos genotipados en cada gen.

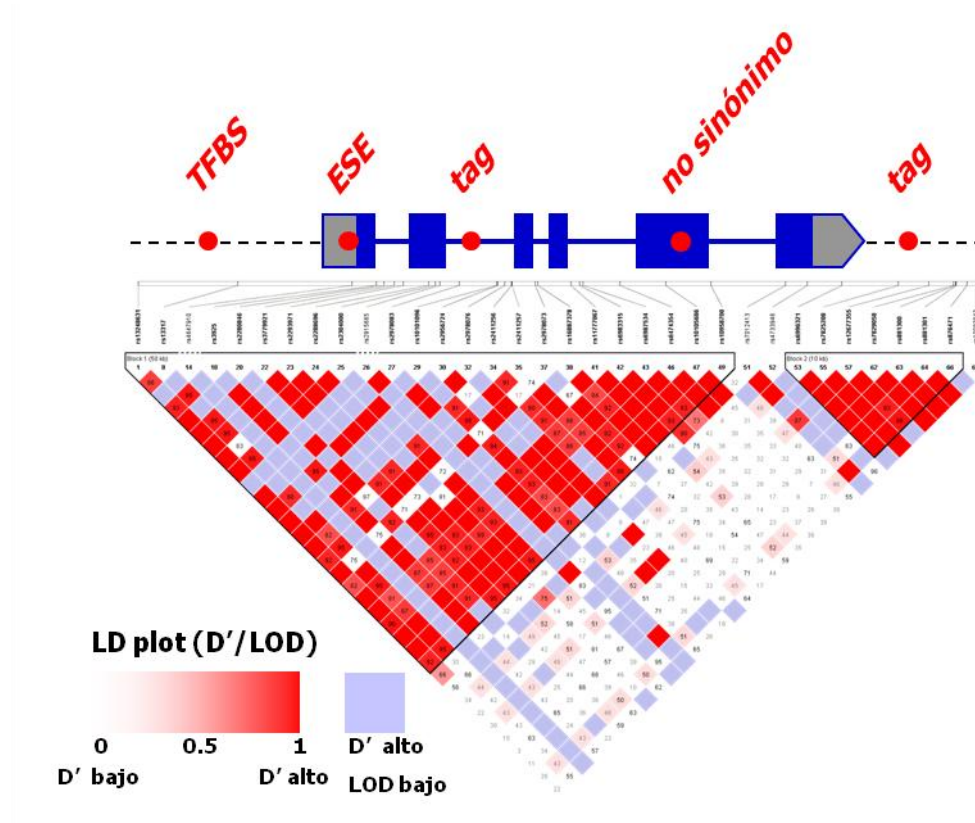


Figura M3. Gen, tipos de SNPs y LD. Representación esquemática de los cuatro tipos de SNPs genotipados en el estudio, y del patrón de disequilibrio de ligamiento observado para un gen con el programa Haploview.

3.2.2. Criterios adicionales de selección de SNPs

Dado que el número de SNPs susceptibles de ser genotipados fue muy alto (varios miles, sólo para representar los genes candidatos elegidos), tuvimos que emplear una serie de criterios adicionales para reducir el número de polimorfismos, tratando, al mismo tiempo, de no perder información sobre los genes seleccionados. Los filtros aplicados para la reducción del número de SNPs fueron:

1. Criterios relativos a la frecuencia alélica del alelo minoritario o MAF (*Minor Allele Frequency*). Se estableció como umbral un MAF mínimo del 10% para tagSNPs y del 2% para SNPs potencialmente funcionales.
2. Criterios relativos a parámetros técnicos de la plataforma de genotipado. El *Illumina score* es un valor que cuantifica la probabilidad de éxito en el genotipado de un SNP mediante la plataforma de elección, la *Illumina*

GoldenGate (apartado 4.1). Se basa en varias fuentes de error posibles, como la presencia de regiones repetitivas o una temperatura de hibridación incompatible. Sólo se seleccionaron polimorfismos con un *score* superior a 0.6 (lo recomendado por el fabricante).

De esta forma, el número final de polimorfismos para genotipar fue de 768 SNPs. Como puede observarse en el orden de genes de la tabla M2, los 63 mejores genes candidatos fueron representados con tagSNPs y SNPs funcionales, y los 34 genes restantes, sólo con variantes no sinónimas. Además, dentro de los 63 genes mencionados, para 7 de ellos (*JAK1*, *GHR*, *CSGALNACT1*, *THADA*, *TPO*, *TG* y *TSHR*), dada su extensión, sólo se eligieron SNPs funcionales y tagSNPs que capturasen la variación de esos bloques de LD. El número de SNPs seleccionados en cada gen, y la categoría a la que pertenecen, está más detallada en la tabla S1.

Por último, referido al gen *CDKN1B* (p27), adoptamos una aproximación directa, dada la ausencia de bloques de LD en la región, eligiendo, por tanto, sólo SNPs funcionales. De los 15 SNPs descritos en este *locus*, sólo 8 presentaron un MAF suficiente en población caucásica, y de ellos, tan sólo tres, alguna predicción de funcionalidad. De esta forma, seleccionamos para el genotipado las variantes rs36228499 (-838C>A), rs34330 (c.-79C>T) y rs2066827 (c.326T>G, p.V109G).

4. Estrategias de genotipado de polimorfismos

En este estudio empleamos distintas plataformas de genotipado, en función del número de muestras y SNPs a genotipar. En general, empleamos el sistema Illumina GoldenGate® (apartado 4.1) para la fase de descubrimiento del estudio (serie 1 de casos y controles), ya que permite el estudio simultáneo de muchas variantes en cada ensayo. Para la fase de validación (series 2 y 3), empleamos la plataforma KASPar (apartado 4.3), y para el estudio de variantes en *CDKN1B*, principalmente la plataforma TaqMan (apartado 4.2), dado que estos dos últimos sistemas permiten analizar un gran número de muestras para una variante concreta.

4.1. Plataforma de genotipado Illumina GoldenGate®

4.1.1. Descripción de la plataforma

El sistema de genotipado empleado en la fase de descubrimiento de nuestro estudio (serie 1) fue la plataforma *Illumina GoldenGate® Genotyping Assay* (San Diego, CA, USA) que nos permitió el estudio simultáneo de 768 polimorfismos. Se empleó como soporte físico una matriz universal *Sentrix-96 Array* (Fig M4A), lo cual nos permitió el estudio de los 768 SNPs en 96 muestras en cada ensayo. Las placas utilizadas contenían matrices de 8 x 12, cada una de ellas realizada en un material conductor de luz (Figs M4B y C). Cada matriz cuenta con 50000 puntos de genotipado (Fig M4D), de forma que se obtuvieron, de media, 65 análisis genotípicos por *locus* (50000 puntos/768 SNPs). Cada punto de genotipado, a su vez, está formado por una fibra óptica (Fig M4E), químicamente modificada para crear un pocillo microscópico al final de cada una de ellas. En este pocillo se inserta una cuenta, que contiene a su vez múltiples copias de un único oligonucleótido o "secuencia señal", que corresponde a una única variante.

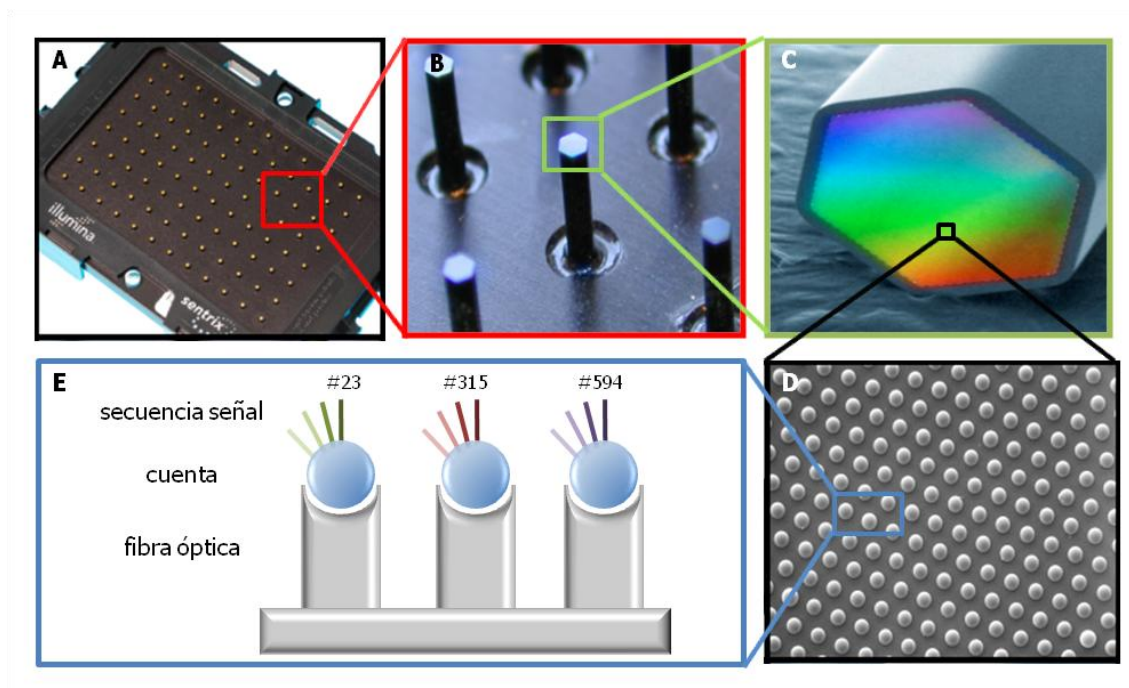


Figura M4. Plataforma de genotipado Illumina Sentrix® Array. **A.** Placa con 96 matrices, cada una de ellos conteniendo 50000 puntos de genotipado; **B.** Paquete de fibra óptica; **C.** Imagen ampliada del paquete que contiene las 50000 fibras ópticas; **D.** Imagen microscópica que muestra los puntos de genotipado que conforman cada una de las matrices de la plataforma; **E.** Esquema de los puntos de genotipado del array, en los que cada cuenta, insertada en un pocillo de fibra óptica, contiene múltiples copias de la misma secuencia señal, cada una de las cuales corresponderá con un SNP del ensayo.

4.1.2 Principios metodológicos del sistema Illumina GoldenGate

El proceso de genotipado mediante el sistema Illumina GoldenGate sigue un protocolo específico y altamente automatizado que dura dos días y medio por placa. Una descripción detallada de esta tecnología puede encontrarse en la web del proveedor (http://www.illumina.com/technology/goldengate_genotyping_assay.ilmn) y en la revisión de Shen y colaboradores (Shen, et al., 2005).

Los pasos principales del protocolo de genotipado, resumidos en la figura M5, consisten en:

1. La activación del ADN, que supone la incorporación del ADNg en forma de cadena sencilla a un soporte sólido, constituido por bolas paramagnéticas. Esto permitirá posteriormente la captura del ADN y la eliminación de impurezas y

oligonucleótidos no hibridados correctamente, disminuyendo el ruido de forma considerable.

2. La discriminación alélica mediante esta plataforma se basa en la extensión alelo-específica de las regiones de interés, para la que se hibridan dos tipos de oligonucleótidos:

- Un oligonucleótido específico de *locus* (P3 en la figura M5A), que consta de una secuencia específica de *locus* y una región que actúa como "secuencia señal", al ser complementaria a la existente en cada uno de los 768 diferentes puntos de genotipado. Además, consta de una secuencia complementaria para un cebador universal.
- Dos oligonucleótidos específicos de los alelos de interés (P1 y P2 en la figura M5A), que difieren en su última base (en 3'), complementaria a cada uno de los dos alelos del SNP de interés, mientras que la región 5' porta secuencias complementarias a dos cebadores universales.

La especificidad de esta extensión es muy alta, ya que se requiere la hibridación de uno de los dos cebadores sentido (P1 ó P2) y del antisentido (P3). Además, el ensayo emplea una polimerasa con alta especificidad por el extremo 3' y sin actividad exonucleasa. De esta forma, la enzima añade nucleótidos para rellenar el hueco entre P1/2 y P3, y una ligasa sella el producto (figura M5B), de forma que los amplicones incompletos no pasarán a la siguiente fase. El hecho de que esta extensión alelo-específica se realice previamente a la PCR evita el posible sesgo que toda amplificación puede producir.

3. El resultado del paso anterior es la obtención de secuencias de ADN molde (en nuestro caso, 768 diferentes), que contienen el SNP de interés, la secuencia señal para su posterior hibridación en el array, y dos secuencias complementarias a cebadores universales (figura M5C), de los que se emplean dos clases:

- Dos cebadores, marcados cada uno de ellos con un fluorocromo (Cy3 ó Cy5), que constan de una secuencia complementaria a los oligonucleótidos alelo-específicos P1 y P2.
- Un tercer cebador complementario a la zona P3 del oligonucleótido específico de *locus*, que además está biotinilado.

Con estos elementos, se realiza una amplificación mediante una PCR convencional, con la particularidad de que en cada pocillo de la placa de 96, correspondiente a una muestra, se generarán 768 productos diferentes, uno por cada SNP estudiado. Las condiciones de amplificación son:

- Pasos iniciales: 37 °C, 10 min; 95 °C, 3 min
- 35 ciclos de desnaturalización, hibridación y extensión: 95 °C, 35 s; 56 °C, 35 s; 72 °C, 2 min.
- Pasos finales: 72 °C, 10 min; 4 °C, 5 min.

4. Los productos de PCR se purifican, gracias a la biotina añadida a los amplicones durante la PCR, y se procede a su hibridación con el *Sentrix-96 Array*, de forma que cada producto, que porta un fluorocromo, se unirá específicamente a cada cuenta del array, gracias al reconocimiento de ambas secuencias señal (figura M5D). Este paso se realiza en un horno de hibridación, primero a 60 °C durante 30 min, y luego a 45 °C durante toda la noche (al menos 14 h).

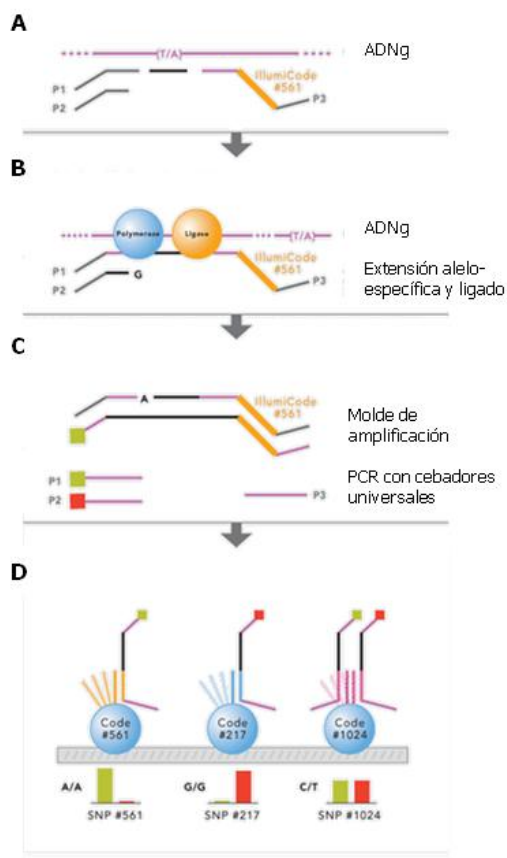


Figura M5. Esquema del proceso de genotipado Illumina GoldenGate®. **A.** Hibridación de oligonucleótidos (P1 ó P2, y P3) al ADN genómico activado; **B.** Extensión alelo-específica, gracias a los oligonucleótidos específicos de SNP (P1 ó P2) y *locus* (P3), y posterior ligado de productos; **C.** Amplificación por PCR con cebadores universales, marcados con fluorocromos; **D.** Hibridación de productos fluorescentes a la matriz *Sentrix-96 Array*, donde las secuencias señal se reconocen específicamente, y la fluorescencia alelo-específica permite la correcta discriminación de los genotipos de cada SNP.

4.1.3. Interpretación de los resultados

Cada array, que contiene la información de 768 SNPs para 96 muestras, se lee en un escáner *BeadArray Reader* de Illumina. Este sistema emplea un láser para excitar los fluorocromos Cy3 y Cy5 del producto de PCR unido a las cuentas de la fibra óptica (fig M4E array págs. anteriores). La emisión de estos fluorocromos (a una longitud de onda de 532 nm para Cy3, y de 635 nm para Cy5), se graba en imágenes de alta resolución, que muestran los miles de puntos de genotipado para cada muestra (fig M6A). El escáner convierte las intensidades de fluorescencia en tres colores: verde y rojo, correspondientes a muestras homocigotas, y amarillo, combinación de las dos señales anteriores, y que representa a los heterocigotos.

Las señales de fluorescencia obtenidas para cada SNP se interpretan utilizando el programa *BeadStudio*, versión 3.1 de Illumina. Este software permite, como se muestra en la figura M6B y M6C, la distinción de los tres grupos de genotipos posibles: homocigotos salvajes (*wildtype*), heterocigotos y homocigotos polimórficos. Aunque la asignación de genotipos en función de su fluorescencia es automática, el programa permite también determinar manualmente aquéllas muestras no clasificadas de forma correcta. Finalmente, los genotipos se exportan a un archivo de texto sin formato (*.csv) para su análisis posterior.

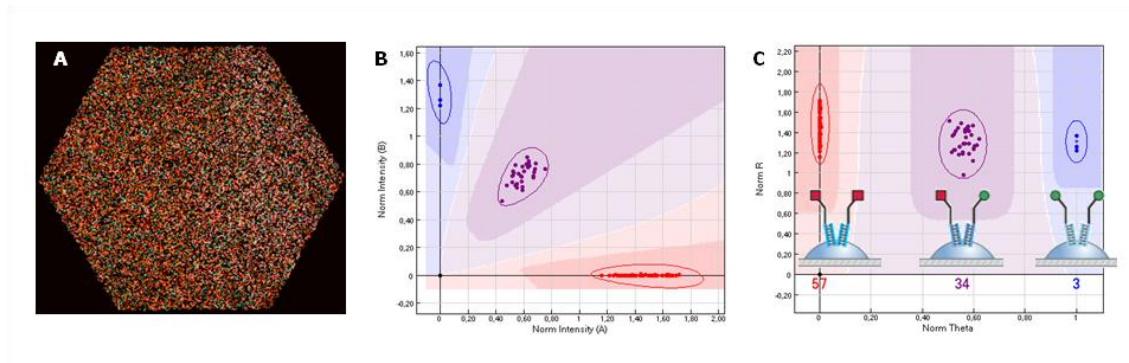


Figura M6. Análisis de los resultados del genotipado Illumina GoldenGate. **A.** Imagen real de la fluorescencia obtenida para una muestra de nuestro ensayo. El hexágono incluye los 50000 puntos de genotipado.; **B.** Análisis de los resultados mediante el programa *BeadStudio*. La imagen muestra la representación cartesiana de los grupos de genotipos de un SNP para una placa de 96 muestras; **C.** Representación polar de los mismos resultados, y esquema de las señales de fluorescencia que los determinan.

4.1.4. Protocolo específico de genotipado

Para nuestro ensayo, utilizamos una cantidad de ADN_g inicial de 400 ng, sobrepasando así los requisitos mínimos de la plataforma, que teóricamente funciona con 250 ng. Cada placa de 96 pocillos se preparó y genotipó por separado, hasta un total de 13 placas. En cada placa se incluyó un blanco, un duplicado intraensayo y un duplicado interensayo, obteniendo un 100 % de concordancia. Además, todas las placas incluyeron casos y controles para ser genotipados en el mismo ensayo.

Para determinar la especificidad del genotipado y facilitar la asignación inicial de genotipos a cada SNP, en la primera placa se incluyeron 9 muestras de genotipos conocidos, consistentes en tres tríos (padre-madre-hijo) de origen caucásico incluidas en el Proyecto HapMap. Los códigos de las muestras de la población HapMap-CEU incluidas fueron: NA10835, NA10860, NA10861, NA11992, NA11993, NA11994, NA11995, NA12248 y NA12249. La segregación de los 768 SNPs en estos tríos fue correcta en todos los casos, y el patrón de separación de los tres genotipos en la placa 1 se exportó al resto de placas, facilitando así una discriminación alélica correcta y homogénea.

4.2. Plataforma de genotipado TaqMan®

4.2.1. Principios metodológicos del genotipado mediante sondas TaqMan®

El genotipado de polimorfismos de un único nucleótido mediante la plataforma TaqMan® (Applied Biosystems, CA, USA) se basa en la detección directa de un producto de PCR como consecuencia de la liberación de un fluorocromo. Para ello, se aprovecha la actividad exonucleasa 5'-3' de la enzima ADN polimerasa, así como el principio físico conocido como "transferencia de energía por resonancia de fluorescencia" (FRET).

En el ensayo se emplea una pareja de cebadores no marcados, que permiten la amplificación de la región de interés, y dos sondas marcadas, cada una de ellas específicamente diseñadas para discriminar cada uno de los dos alelos de la variante. La detección específica de cada alelo se consigue gracias a dos tipos de moléculas unidas covalentemente a cada sonda TaqMan® (Fig M7):

- En el extremo 5', una molécula fluorogénica o fluorocromo, normalmente FAM o VIC, cada una de ellas unida a la sonda para cada uno de los alelos.
- En el extremo 3', un grupo receptor de fluorescencia o *quencher*, cuya longitud de excitación coincide con la longitud de onda de emisión del fluorocromo del extremo 5' (FAM o VIC). De esta forma, la proximidad del *quencher* al fluorocromo inhibe la emisión fluorescencia de este último.

El genotipado consiste en una reacción de PCR convencional, con la particularidad de que, además de los cebadores, también se hibridan la(s) sonda(s), en el caso de que su alelo complementario esté presente en la secuencia problema. Como se muestra en la figura M7, tras la desnaturalización del ADN y la hibridación de los cebadores, la polimerasa comienza a sintetizar la hebra correspondiente. Al llegar a la región donde se encuentra la sonda, la degrada mediante su actividad exonucleasa 5'→3'. Esta degradación hace que el fluorocromo se aleje del *quencher*, emitiendo fluorescencia en cada ciclo de amplificación y degradación de la sonda. Así, al final del proceso, las muestras heterocigotas generarán señal fluorescente para ambos fluorocromos (VIC y FAM), mientras que las muestras homocigotas producirán señal de sólo uno de ellos.

Cada reacción de genotipado TaqMan®, que se realiza en placas de 384 pocillos, tiene un volumen total de 5 µl, y consta de:

- Una cantidad de ADN problema que puede variar entre los 5 y los 15 ng. Nosotros utilizamos 2.25 µl de ADN a una concentración de 5 ng/µl.
- La solución *TaqMan Universal PCR Master Mix*, que contiene la ADN polimerasa empleada en esta reacción y su tampón de reacción.
- Una mezcla que contiene los cebadores no marcados, así como las sondas TaqMan (marcadas con FAM y VIC) para la detección de los 2 alelos. Todos estos componentes están diseñados específicamente para cada uno de los ensayos de discriminación alélica.

Las reacciones de PCR para el genotipado mediante TaqMan tienen las siguientes condiciones estándar:

- Un paso inicial de desnaturalización a 95 °C durante 10 min.
- 40 ciclos de desnaturalización/extensión: 92 °C durante 10 s, seguidos de 1 min a 60 °C.

4.2.2. Protocolo específico de genotipado

El genotipado de los polimorfismos rs34330 y rs2066827 del gen *CDKN1B* se realizó mediante esta tecnología. La capacidad de discriminación de la sonda se probó primero con 24 controles en duplicado. Además, se controló la especificidad del genotipado incluyendo dos duplicados y un blanco por cada placa de 96 pocillos incluida en el estudio. Las reacciones de genotipado se realizaron en placas de 384 pocillos, a las que se dispensaron previamente las muestras desde placas de 96 utilizando un robot *Beckman Coulter Biomek® NX^P* (Brea, CA, USA). En el mismo ensayo se genotiparon casos y controles, de forma que cualquier posible sesgo afectase por igual a ambos tipos de muestras. Por último, se secuenciaron un 5 % de las muestras incluidas en el estudio, obteniendo un 100 % de concordancia con los genotipos discriminados mediante TaqMan®.

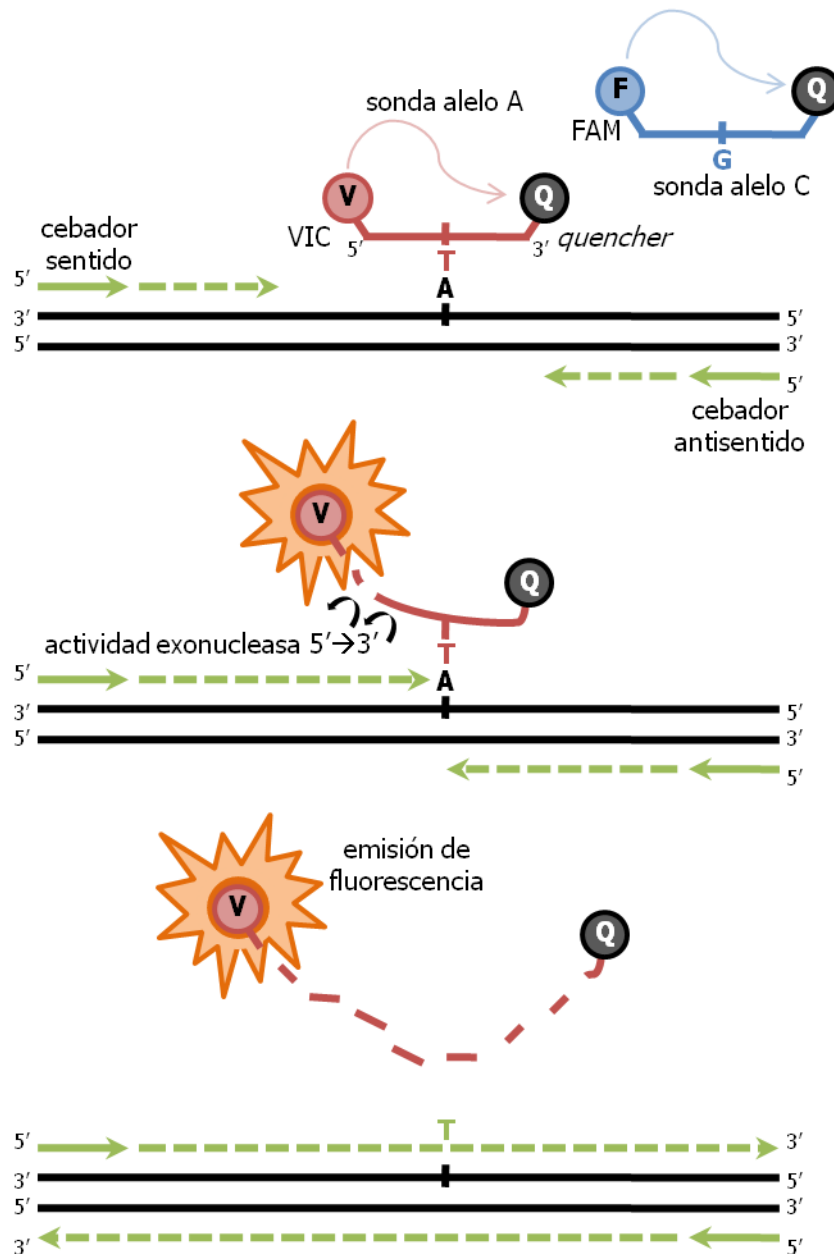


Figura M7. Representación esquemática del genotipado mediante sondas TaqMan®. Las sondas se unen al ADN molde sólo en presencia de su alelo complementario. La actividad exonucleasa de la ADN polimerasa permite el alejamiento del fluorocromo del *quencher*, y la emisión del fluorescencia del primero.

4.3. Plataforma de genotipado KASPar

4.3.1. Principios metodológicos del genotipado mediante sondas KASPar

El sistema de genotipado *KASPar SNP Genotyping System* (Kbioscience, Herts, UK) es un método capaz de discriminar SNPs mediante una reacción de PCR competitiva alelo-específica. Al igual que la plataforma TaqMan[®], utiliza el principio de “transferencia de energía por resonancia de fluorescencia” (FRET), pero a diferencia de ésta, es la actividad polimerasa, y no la actividad exonucleasa, la que produce la emisión de fluorescencia. De hecho, el ensayo KASPar utiliza una enzima Taq polimerasa modificada, que carece de actividad exonucleasa. Los componentes que intervienen en el genotipado se muestran en la figura M8, y consisten en:

- Dos cebadores sentido sin marcar, que compiten por la unión al ADN molde. Cada uno de ellos tiene en 3' la base complementaria a cada uno de los alelos del SNP a genotipar, respectivamente, de forma que sólo se unirán al ADN problema en presencia de la base correspondiente. Además, en 5', estos cebadores portan sendas colas de oligonucleótidos universales, con secuencia desconocida y propiedad de KASPar, que permiten la hibridación posterior con cebadores universales marcados.
- Un cebador antisentido común sin marcar, que confiere especificidad de *locus*, pero que se une en presencia de ambos alelos del SNP.
- Dos cebadores universales marcados, cada uno de ellos con una secuencia complementaria a las colas de los cebadores sentido. Estos cebadores tienen unidas dos moléculas: un fluorocromo (FAM o VIC) y un grupo receptor de fluorescencia o *quencher*. La estructura de horquilla de esta molécula hace fluorocromo y *quencher* estén muy próximos, de forma que no hay emisión de fluorescencia detectable.

Como se muestra en la figura M8, el proceso de genotipado comienza con la hibridación competitiva de los cebadores en función de la variante de interés. Durante la fase de extensión, se añaden las colas universales a los amplicones de la región del SNP. En los siguientes ciclos, los cebadores universales marcados serán capaces de unirse a esas colas, sustituyendo a los cebadores iniciales. El alejamiento del fluorocromo del *quencher* produce emisión de fluorescencia detectable a la longitud de onda de FAM, VIC o ambos, que indicará la presencia de un alelo, otro o los dos, respectivamente.

Cada reacción de genotipado KASPar se realiza sobre un volumen total de 4 µl, y consta de:

- El ADN genómico problema. Utilizamos 15 ng de ADNg, dispensando 3 µl de una dilución a 5 ng/µl.
- La solución *KASPar reaction mix*, que contiene la polimerasa modificada, su cofactor MgCl₂, el tampón de reacción y los cebadores universales marcados con fluorescencia.
- El reactivo específico de cada reacción, con los cebadores sentido y antisentido diseñados por Kbioscience para amplificar la región de interés.
- En algunos casos, la reacción se suplementa con MgCl₂ adicional, hasta una concentración final de 2.2 mM.

La secuencia de los cebadores utilizados y las condiciones de PCR empleadas, están resumidas en la tabla M3.

4.3.2. Protocolo específico de genotipado

El sistema de genotipado KASPar fue la plataforma elegida para la fase de validación de nuestro estudio de asociación (series 2 y 3), ya que permite el genotipado de un alto número de muestras a un coste menor. En total, genotipamos 23 polimorfismos con esta tecnología (tabla M3). Utilizamos un robot *Beckman Coulter Biomek[®] NX^P* (Brea, CA, USA) para dispensar 15 ng de ADNg de cada muestra desde placas de 96 pocillos a placas de 384, donde se realizaron las reacciones de genotipado. La capacidad de discriminación de cada sonda, se probó primero en 24 controles dispensados en duplicado. Además, ya con las muestras de interés, por cada placa de 96 pocillos, se incluyeron dos duplicados (uno intraplaca y otro interplaca) y dos blancos, obteniéndose un 100 % de concordancia. Por último, siempre se genotiparon en la misma placa de 384 tanto casos como controles, para minimizar los un posible sesgo interensayo.

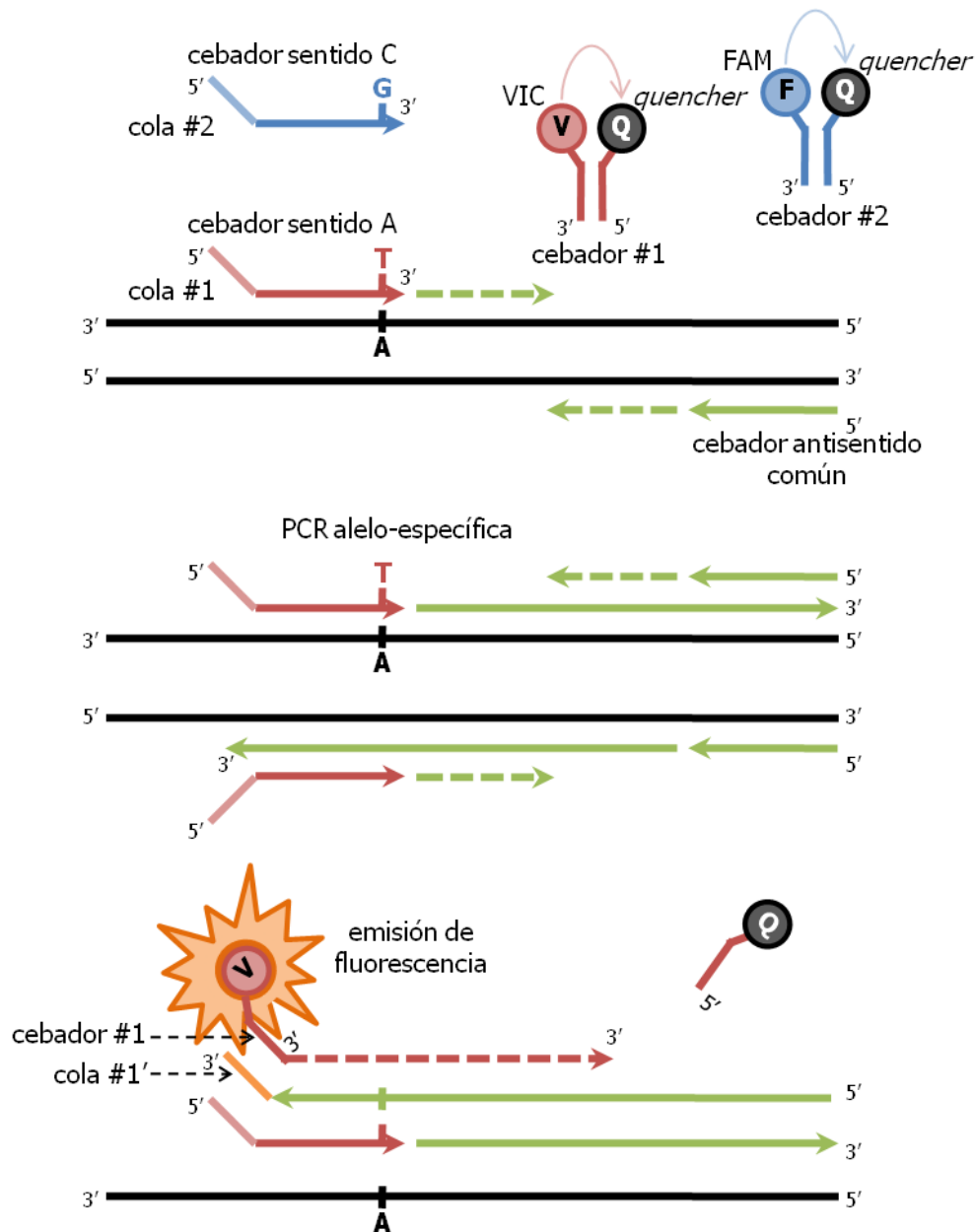


Figura M8. Representación esquemática del genotipado mediante el sistema KASPar. Inicialmente, los cebadores sentido compiten por la unión al ADN en función de la secuencia del SNP. Además, se añaden colas universales a los amplicones generados. Estas colas son reconocidas por los cebadores marcados, cuyo alejamiento del *quencher* permite la emisión de fluorescencia. Adaptado a partir de lo descrito en la web del proveedor de la tecnología KASPar: <http://www.kbioscience.co.uk/download/KASP.swf>

Tabla M3. Polimorfismos estudiados mediante sondas KASPar

Variante, alelos	Gen	Cebador sentido, alelo 1 (FAM)	Cebador sentido, alelo 2 (VIC)	Cebador antisentido común	Condiciones PCR (ciclos, T _h , [MgCl ₂])	Estudio
rs1867277, G>A	<i>FOXE1</i>	CCAGAGTCCAGTCCCG GTCA	CAGAGTCCAGTCCCGG TCG	TKGTGCTTCTCGAGG CGGGCA	40c, 61 °C, 2.2mM	Individual, Epistasia
rs2284734, A>G	<i>TSHR</i>	CCTAAGTATGTAAGTT ACTGACACACA	CTAAGTATGTAAGTTA CTGACACACG	GAAGACACTTCTGCA TGTATAGTGCAATA	38c, 57 °C, 2.2mM	Individual, Epistasia
rs2687834, G>T	<i>TG</i>	CCTTCTGAAGAATAAG ACTCCCTTC	CCTTCTGAAGAATAAG ACTCCCTTA	GGTATGACCATGAGA CAGAGAGCTT	38c, 57 °C, 2.2mM	Individual
rs25882, T>C	<i>CSF2</i>	GAAACTTCCTGTGCAA CCCAGAC	GAAACTTCCTGTGCAA CCCAGAT	CAGGTTCTCTTTGAA ACTTTCAAAGGTGAT	38c, 57 °C, 2.2mM	Individual
rs2703488, T>C	<i>KIT</i>	AACAAAGAGATGATAC TGGAAATAGCG	CAAAACAAAGAGATGAT ACTGGAAATAGCA	CTCATACTCCCCCTCC CCCAT	38c, 57 °C, 2.2mM	Individual, Epistasia
rs13099828, C>G	<i>PPARG</i>	CTTAATATTTAAATTTG AATCGGAGCTTG	CTCTTAATATTTAAAT TTGAATCGGAGCTTC	CTCCCCAGATAATTTT AATATGTAGCCAAA	38c, 57 °C, 2.2mM	Individual
rs16973034, A>G	<i>PRKAR1A</i>	CCTTAGCGTTGTGAGA GAAAATGCAT	CTTAGCGTTGTGAGAG AAAATGCAC	GCTCTAATAATGGTG ATGCTGGGCTA	38c, 57 °C, 2.2mM	Individual
rs6179, G>A	<i>GHR</i>	AGTTCATACTCCAGAA CCATCCAT	AGTTCATACTCCAGAA CCATCCAC	GGAAGCACCACGCAA TGCAGATATT	38c, 57 °C, 2.2mM	Individual
rs2066807, G>C	<i>STAT2</i>	CGCAGTAGAAAGGTGC CAGAC	CGCAGTAGAAAGGTG CCAGAG	CAGGAGCGCCGGCTG CTGAA	38c, 57 °C, 2.2mM	Individual, Epistasia
rs1053266, G>T	<i>CCDC6</i>	GCATTAAGGCTGGGAG GAGGG	GCATTAAGGCTGGGA GGAGGT	CGCAGCCTACTCCTT CGCAACAT	38c, 57 °C, 2.2mM	Individual
rs4939827, T>C	<i>SMAD7</i>	GAGGGAGCTCTGGGGT CCTA	AGGGAGCTCTGGGGT CCTG	CTCACAGCCTCATCC AAAAGAGGAA	38c, 57 °C, 1.8mM	Individual, Epistasia
rs4848323, C>T	<i>PAX8</i>	GCCCCACTGATTGATA AGAAGTAG	GCCCCACTGATTGATA AGAAGTAA	GCTTCCAATAGACAT CAAAAAGTGTCAAT	38c, 60 °C, 1.8mM	Epistasia
rs1378624, G>A	<i>STK17B</i>	AGATTACCATATTTACC ATTACTTATCTACA	ACCATATTTACCATTA CTTATCTACG	CTCAAAGAGTTTTGG TTGATGCAGGTTAT	38c, 60 °C, 1.8mM	Epistasia

Variante, alelos	Gen	Cebador sentido, alelo 1 (FAM)	Cebador sentido, alelo 2 (VIC)	Cebador antisentido común	Condiciones PCR (ciclos, T _h , [MgCl ₂])	Estudio
rs11705805, A>G	<i>RAF1</i>	CAGATTCTTTATCTGTT TAAATTTAAAGTCCA	CAGATTCTTTATCTGT TTAAATTTAAAGTCCG	CTGAACTCTGAACTC TGAATGTTATGGTA	38c, 60 °C, 1.8mM	Espistasia
rs6554198, A>G	<i>KIT</i>	AATAGGGAGTTATACC ATGTTGGACT	ATAGGGAGTTATACCA TGTTGGACC	CCTAGGTGCTCAAAA GATATTTGCTGAAT	38c, 60 °C, 1.8mM	Espistasia
rs376618, A>G	<i>FGFR4</i>	GGGGTAACTGTGCCTA TTCGAGA	GGGTAACTGTGCCTAT TCGAGG	ATGATGAGGACCCCA AGTCCCATA	38c, 60 °C, 1.8mM	Espistasia
rs1044141, G>A	<i>STK17A</i>	AACCTCTTCGGTTACAAT GGATTCCTT	ACTCTTCGGTTACAAT GGATTCCTC	GTCATTCTGTGCCTG AAATTAATTCGGAT	38c, 60 °C, 1.8mM	Espistasia
rs7816924, C>T	<i>CSGALNACT1</i>	ACTGCCATCCTGGTAA ATAAACACG	ATACTGCCATCCTGGT AAATAAACACA	GGAGCCTGGCTTCCA GGTTGTT	38c, 60 °C, 1.8mM	Espistasia
rs2712622, T>C	<i>SLC5A8</i>	CCTAGTAATTGATGGC TGAGATCG	GCCTAGTAATTGATGG CTGAGATCA	CAACTCAAGACAGAA CCAAGAAGTCTATT	38c, 60 °C, 1.8mM	Espistasia
rs721992, G>T	<i>CCDC6</i>	ACTGTGAACTCTGAAG AGGTACTC	GACTGTGAACTCTGAA GAGGTACTA	TGTTTCTGAGAGCAT GCCTAGCCAA	38c, 60 °C, 1.8mM	Espistasia
rs25879, A>G	<i>CSF2</i>	GATTCTAACTCCTCTTA CTCACATTCAA	CTAACTCCTCTTACTC ACATTCAG	CCCCTCCCCCCCCAC CAAA	38c, 60 °C, 1.8mM	Espistasia
rs4422036, G>A	<i>MED13</i>	TACTCATTTTGTAGCA GTAAGTCAATACT	CTCATTTTGTAGCAGT AAGTCAATACC	ATCCAGCCTTTGCAA ATTACTTTGCTCAA	38c, 60 °C, 1.8mM	Espistasia
rs2979033, G>A	<i>TG</i>	TGCTTTTATTTCTATGC TAATTCCTTCATAT	GCTTTTATTTCTATGC TAATTCCTTCATAC	GACAGCCACTGGAAC GAACTATTATATTT	38c, 60 °C, 1.8mM	Espistasia

Las colas universales en 5' de los cebadores sentido no se muestran, por ser propiedad de *Kbioscience*. Abreviaturas: T_h=temperatura de hibridación; mM=milimolar; [MgCl₂]=concentración de cloruro de magnesio

4.3.3. Interpretación de los resultados de genotipado por TaqMan y KASPar

El análisis del genotipado mediante TaqMan y KASPar se realiza utilizando el mismo equipo. La fluorescencia a tiempo final, que es el resultado de la cantidad de ADN inicial y, sobre todo, de la presencia de un determinado alelo de la variante a estudiar, se detectó utilizando el equipo *ABI PRISM® 7900 HT Sequence Detector* (Applied Biosystems, Foster City, CA, USA). Después, mediante el programa SDS, versión 2.2.2, suministrado por la misma casa comercial, se analizaron los distintos grupos de fluorescencia, como se muestra en la figura M9. Para el análisis de los tres grupos para cada SNP (homocigotos para el alelo *wildtype*, heterocigotos y homocigotos del alelo polimórfico), siempre se tuvo en cuenta el patrón de fluorescencia (intensidad y capacidad de discriminación) obtenido en el grupo inicial de controles en el que se probó cada sonda, de forma que, aquellas muestras para las que el programa no asignó un genotipo automáticamente, se agruparon de forma manual si reunían las condiciones adecuadas. Finalmente, los genotipos se exportaron, ya en formato de texto ("a1a1", "a2a2", "both", "undetermined"), primero a un archivo de texto sin formato (*.txt), y luego a una tabla Excel, para su posterior conversión en los genotipos correspondientes y análisis estadístico.

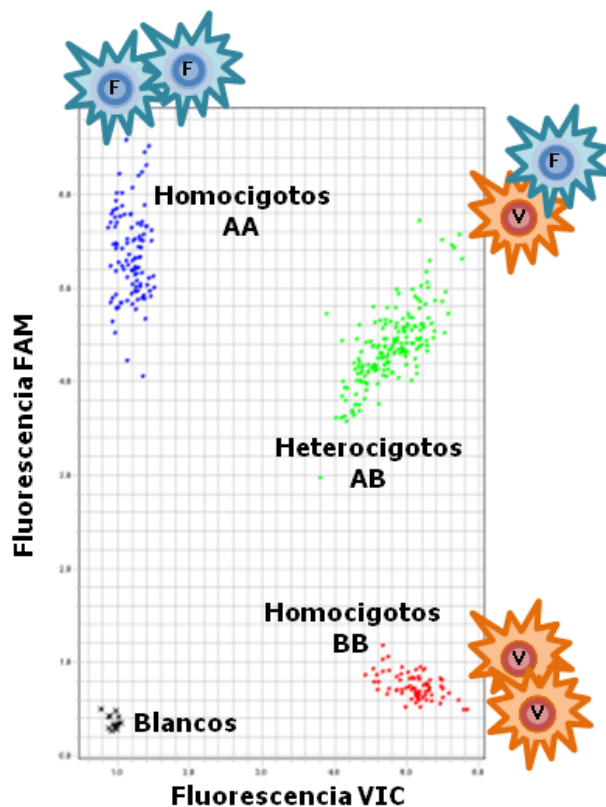


Figura M9. Análisis de los genotipos obtenidos con las plataformas TaqMan o KASPar. La imagen muestra la asignación automática de genotipos que realiza el programa SDS 2.2.2, en función de los valores de fluorescencia obtenidos para cada muestra, representadas como puntos de colores en los ejes cartesianos.

4.4. Genotipado por digestión diferencial de fragmentos (RFLP)

El SNP rs36228499 (-838C>A) del gen *CDKN1B* (p27), se genotipó mediante una técnica de digestión diferencial o RFLP (*Restriction Fragment Length Polymorphism*), de forma similar a lo descrito por González y colaboradores (Gonzalez, et al., 2004). Para mejorar la intensidad de la amplificación y, por tanto, la resolución de la técnica, modificamos ligeramente el protocolo, diseñando una segunda PCR a partir de la primera, utilizando para ello cebadores internos. Los cebadores utilizados y las temperaturas de hibridación o *annealing* (T_h) utilizadas fueron las siguientes: 1ª PCR, 838_S: 5'-GACCTTCGCGGTCCTCTG-3'; -838_AS: 5'-GTGACTGCTGGAGGGGTACT-3'; T_h =58 °C, 40 ciclos; 2ª PCR: -838N_S: 5'-TCCAGGTCCCGGCTTCCCGGt-3'; -838N_AS: 5'-CCTGCTCTGGCTGGCCTCGGAG-3'; T_h =65 °C, 35 ciclos. El cebador -838N_S incluye un desemparejamiento (mostrado en minúscula en el cebador) que permite crear una secuencia de reconocimiento para la endonucleasa de restricción *TaqI*, sólo cuando el alelo C del SNP rs36228499 está presente.

Para la digestión, se emplearon 20 µl del producto de la segunda PCR, se añadió 0.05 µl de la enzima *TaqI* (Fermentas, Burlington, Canada) por muestra, y se incubó durante toda la noche a 65 °C. Los productos resultantes se corrieron en un gel de agarosa D1 medium (Pronadisa, España) al 3.5 % durante 90 min. La digestión enzimática, como se muestra en la figura M10, produjo fragmentos fácilmente distinguibles en función del genotipo del SNP: 173 bp (no digerido, corresponde al genotipo polimórfico AA), 152+21 bp (digerido, genotipo salvaje CC) o ambos (heterocigotos CA).

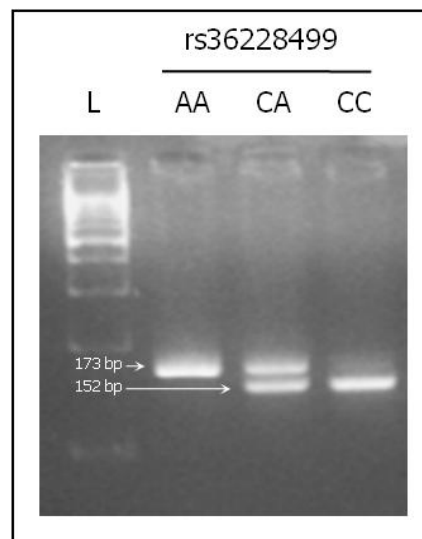


Figura M10. Genotipos representativos del rs36228499, discriminados por RFLP

4.5. Genotipado mediante DHPLC

El DHPLC (*"Denaturing High Performance Liquid Chromatography"*) es un método basado en la detección de especies heteroduplex. Los heteroduplex se forman mediante la desnaturalización por calor del ADN y posterior renaturalización del mismo, lo cual produce un desemparejamiento en individuos heterocigotos. Estos heteroduplex son distinguibles de los homoduplex, formados por las dos cadenas idénticas presentes en individuos homocigotos.

El DHPLC se basa en la utilización de una columna compuesta por moléculas hidrofóbicas, que en fase reversa unirá especies de ADN mediante TEEA (trietilamonio-acetato). El ADN es eluido de la columna mediante la incorporación del compuesto hidrofóbico acetonitrilo, en un porcentaje de acuerdo a la longitud y composición GC del fragmento a estudiar. La adecuada temperatura de desnaturalización parcial del fragmento nos permitirá discriminar especies homoduplex de las heteroduplex por el tiempo que el fragmento esté retenido en la columna (figura M11).

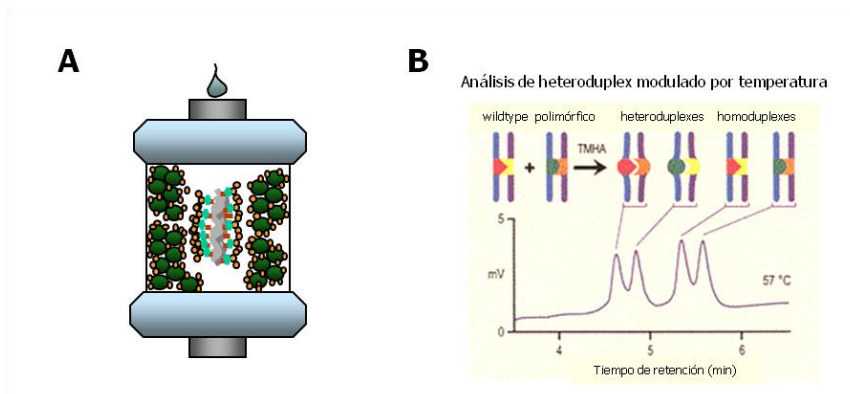


Figura M11. Discriminación alélica mediante DHPLC. **A.** Columna hidrofóbica. En verde, bolas hifrofóbicas; en gris, molécula de doble cadena de ADN; en naranja, TEEA (elemento de unión entre el ADN y los componentes hidrofóbicos de la columna); **B.** Esquema representativo de un cromatograma y su correspondencia con las diferentes especies (homo y heteroduplex). Los picos del cromatograma se distinguen por tiempo de retención

En nuestro trabajo, empleamos el genotipado mediante DHPLC para comprobar experimentalmente, en una serie de 200 controles, el LD total descrito para los SNPs rs907577 y rs1867277 de *FOXE1*. Los patrones más representativos de cada genotipo fueron comprobados mediante secuenciación directa (apartado 4.6).

4.6. Genotipado por secuenciación directa

Para algunas muestras y determinados SNPs, cuyos genotipos no se pudieron discriminar por otras técnicas, se empleó la secuenciación directa para determinar dichos genotipos. Esta técnica también se empleó como confirmación de los resultados obtenidos en otras plataformas de genotipado. Se realizaron reacciones de PCR convencional, en un volumen total de 25 µl, y empleando para la amplificación un termociclador *epgradient S* (Eppendorf). Para el diseño de cebadores, se utilizó el programa web *Primer3* (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). El protocolo general de amplificación fue:

- Desnaturalización inicial: 5 min a 94 °C
- Ciclos de desnaturalización/hibridación/elongación: 30s a 94 °C, 30 s a la temperatura de hibridación (T_h) específica (ver tabla M4), 30 s a 72 °C. Para optimizar el rendimiento de la PCR, se realizaron un total de 35 ciclos, distribuidos en 5 ciclos con una $T_h=x+4$, otros 5 ciclos con una $T_h=x+2$ y 25 ciclos con la T_h óptima para ese fragmento.
- Un paso final de elongación: 10 min a 72 °C

La tabla M4 recoge las variantes secuenciadas, las secuencias de los cebadores y las temperaturas de hibridación utilizadas para cada PCR. Antes de la secuenciación, 5 µl de los productos de PCR se corrieron en un gel de agarosa (Pronadisa, España) al 1.5 % durante 20 min, para comprobar la amplificación. El producto de PCR se purificó, empleando el reactivo ExoSAP-IT™ (GE Healthcare), que contiene una mezcla de las enzimas exonucleasa I y fosfatasa alcalina. Para la purificación, se mezclaron 10 µl de producto de PCR con 4 µl del reactivo, y se incubó durante 15 min a 37 °C, seguido de otros 15 min a 80 °C.

Los amplicones purificados fueron secuenciados de forma automática utilizando terminadores fluorescentes Big Dye™ (Applied Biosystems, CA, USA) y un secuenciador multicapilar ABI modelo 3730xl (Applied Biosystems, CA, USA). Para el análisis de los diagramas de picos resultantes de la secuenciación, empleamos el programa FinchTV, versión 1.4, de la misma casa comercial.

Tabla M4. Polimorfismos estudiados por secuenciación directa

Variante	Gen	Localización, alelos	Cebador sentido (5'→3')	Cebador antisentido (5'→3')	T _h	Uso
rs1867277	<i>FOXE1</i>	5'UTR, c.-283G>A	TAAACTAGCGGGCACCACAG	ACAGAGGCTCGGGAGTGAC	58 °C	Confirmación KASPar y DHPLC
rs36228499	<i>CDKN1B</i>	Promotor, -838C>A	TCCAGGTCCCGGCTTCCCGGT	CCTGCTCTGGCTGGCCTCGGAG	65 °C	Confirmación RFLP
rs34330	<i>CDKN1B</i>	5'UTR, c.-79C>T	CCATTTGATCAGCGGAGACT	CACTCGCACGTTTGACATCT	58 °C	Confirmación TaqMan
rs2066827	<i>CDKN1B</i>	Exón, c.T326G, p.V109G	AAGAGTTAACCCGGGACTT	CTCTCCCAAAGCTAAATCAG	56 °C	Confirmación TaqMan
rs4939827	<i>SMAD7</i>	Intrón, g.27942565T>C	GCCAGGATTCATCGTCTGAT	GAATCCCTCCCCCAAATAGA	60 °C	Confirmación KASPar

Abreviaturas: T_h=temperatura de hibridación; UTR= región no traducida (*untranslated region*).

5. Ensayos funcionales realizados en linfocitos de voluntarios sanos

Durante el desarrollo de este trabajo, realizamos una serie de ensayos funcionales destinados a elucidar el papel funcional de determinados polimorfismos. Para ello, uno de los modelos empleados fueron los linfocitos de sangre periférica extraída de voluntarios sanos, dada la relativa facilidad para su obtención y el hecho de ser células humanas no inmortalizadas. Esta serie de voluntarios sanos, que ascendió a 97 individuos, fue independiente de cualquiera de las series de controles empleadas en las distintas fases del estudio de asociación. De estas células, obtuvimos, en la medida de lo posible, ADNg para el genotipado, ARN para estudios de expresión mediante PCR cuantitativa, y proteína, para estudios de inmunoblot. La disponibilidad de estos tres materiales nos permitió correlacionar los resultados entre el genotipo de variantes polimórficas localizadas en el promotor, como el SNP rs34330 en *CDKN1B*, y su efecto funcional en la expresión del ARN mensajero (ARNm) y proteína.

5.1. Obtención de linfocitos a partir de sangre periférica

El primer paso para la obtención de linfocitos es su separación del resto de componentes que componen la sangre. Para ello, se siguen los siguientes pasos:

1. En un tubo Falcon de 50 ml, se añaden 12 ml de Ficoll Histopaque® (Sigma-Aldrich, St. Louis, MO, USA) y con una pipeta Pasteur se añade la sangre muy lentamente, deslizándola por la pared del tubo, de forma que quede una capa de sangre sobre el Ficoll.
2. Se centrifuga durante 20 min a 1600 rpm para separar las fases, obligando a que la aceleración y el frenado del rotor sea lo más suave posible.
3. Con una pipeta Pasteur, transferir a un tubo Falcon de 15 ml la capa de linfocitos, que aparece como una nube blanquecina entre el plasma (arriba) y el ficoll y eritrocitos (abajo).
4. Lavar los linfocitos, resuspendiéndolos en el tampón PBS, y centrifugando durante 5 min a 2000 rpm (con aceleración y frenado normal). Repetir el lavado 3 veces.

El pellet de linfocitos obtenido tras este procedimiento se procesa de forma diferente según el material que se quiera aislar de él. Si queremos obtener ARN (apartado 5.2.1), se resuspende en 1 ml del reactivo TriReagent (Molecular Research Center, Cincinnati, OH), se transfiere a un tubo eppendorf, y se congela a -80 °C si no se va a procesar inmediatamente. En el caso de que se quieran obtener proteínas (apartado 5.3.1), es esencial disgregar el pellet todo lo posible, mediante pulsos de

vortex y fuerza mecánica sobre el propio pellet, antes de congelarlo directamente a -20 °C, lo cual ayuda, junto a todo lo anterior, a lisar los linfocitos de manera eficiente.

5.2. Ensayos relativos al ARN mensajero

5.2.1. Extracción de ARN

Para el aislamiento de ARN a partir de linfocitos resuspendidos en el reactivo TriReagent se siguen los siguientes pasos:

1. Descongelar la mezcla de TriReagent y linfocitos durante 10 min a temperatura ambiente, y homogeneizarla bien.
2. Añadir 200 µl de cloroformo, mezclar vigorosamente e incubar a temperatura ambiente durante 2-3 min.
3. Centrifugar durante 15 min a 4 °C y 12000 *g* para separar las fases.
4. Transferir la fase acuosa a un tubo eppendorf limpio, añadir 500 µl de isopropanol, mover un poco e incubar 10 min a temperatura ambiente.
5. Precipitar el ARN centrifugando 20 min a 4°C y 12000 *g*.
6. Descartar el sobrenadante y lavar el pellet transparente de ARN con 1 ml de etanol al 75%. Centrifugar 5 min a 4°C y 12000 *g*.
7. Eliminar el etanol, secar el pellet al aire durante 1-2 min, y resuspenderlo en 30 µl de agua dietilpirocarbonato (DEPC).
8. Medir la concentración del ARN obtenido utilizando 1 µl del producto y un espectrofotómetro Nanodrop ND-1000 (NanoDrop Technologies, Wilmington, DE, USA). Congelar el ARN a -80 °C si no se va a utilizar de inmediato.

5.2.2. Retrotranscripción del ARN a ADN complementario

El ADN complementario (ADNc) se obtuvo a partir de ARN utilizando la enzima transcriptasa inversa y siguiendo el siguiente protocolo:

1. Añadir, en hielo, 1 µg de ARN total aislado a un tubo estéril de 0.2 ml, y llevar el volumen a 10.6 µl con agua DEPC.
2. Desnaturalizar la mezcla 2 min a 70 °C y transferirla rápidamente a hielo.
3. Preparar la mezcla para la reacción de retrotranscripción, cuyos reactivos y volúmenes por muestra son:
 - 5X First Strand Buffer* (Invitrogen); 4 µl
 - DTT (0,1 M) (Invitrogen); 2 µl
 - dNTPs (10mM); 1 µl
 - Oligo dT (0,5 µg/µl) 1 µl
 - RNAseOUT* (Ambion); 0.7 µl
 - SuperScript II* (200 unidades (u) /µl) (Invitrogen); 0.7 µl
4. Añadir 9.4 µl de la mezcla de reacción a cada muestra de ARN y mezclar.
5. Incubar durante 90 min a 42 °C.

5.2.3. PCR cuantitativa a tiempo real

Los niveles de ARNm del gen *CDKN1B* (p27) se midieron mediante una PCR cuantitativa a tiempo real. Para ello, se empleó una dilución 1:10 del ADNc de los linfocitos de voluntarios sanos incluidos en el estudio. Se dispensaron 2 µl de estos ADNs en triplicado en placas de 384 pocillos, empleando para ello un robot *Beckman Coulter Biomek® NX^p*, y en estas placas se realizaron las reacciones de PCR cuantitativa, midiendo siempre en la misma placa los niveles de expresión del gen problema (*CDKN1B*, en nuestro caso) y un gen estándar o *housekeeping* (el gen de la β-glucoronidasa o *GUS*), para poder normalizar luego los valores de expresión en función del comportamiento particular de cada muestra.

La expresión de *CDKN1B* se midió empleando una sonda de la librería *Universal ProbeLibrary* (Roche). Estas sondas son oligonucleótidos modificados, conocidos como ácidos nucleicos cerrados (LNA – *Locked Nucleic Acids*), análogos al ADN, pero con un enlace adicional en su anillo de ribosa, lo cual les confiere una mayor capacidad y estabilidad en la hibridación con sus bases complementarias. Las sondas son cadenas de 8 ó 9 nucleótidos LNA, con un fluorocromo FAM unido en su extremo 5' y un *quencher* en 3'. La especificidad en la detección se consigue gracias a la combinación de la sonda con una pareja de cebadores, colocados en dos exones distintos del ADNc del gen de interés, amplificando así una región de éste (normalmente de unos 60 nucleótidos). El principio químico de discriminación es similar al empleado por TaqMan (explicado en el apartado 4.2), utilizando la misma mezcla (*Universal Master Mix*, Applied Biosystems), con la única diferencia de que la fluorescencia se mide en cada ciclo ("tiempo real") en lugar de sólo a tiempo final.

Para la detección de *GUS* se empleó el mismo concepto, pero utilizando una sonda de Sigma-Aldrich. Las secuencias de las sondas, y de las parejas de cebadores correspondientes a cada amplicón, están resumidas en la tabla M5.

Tabla M5. Secuencias de los oligonucleótidos empleados en PCR cuantitativa

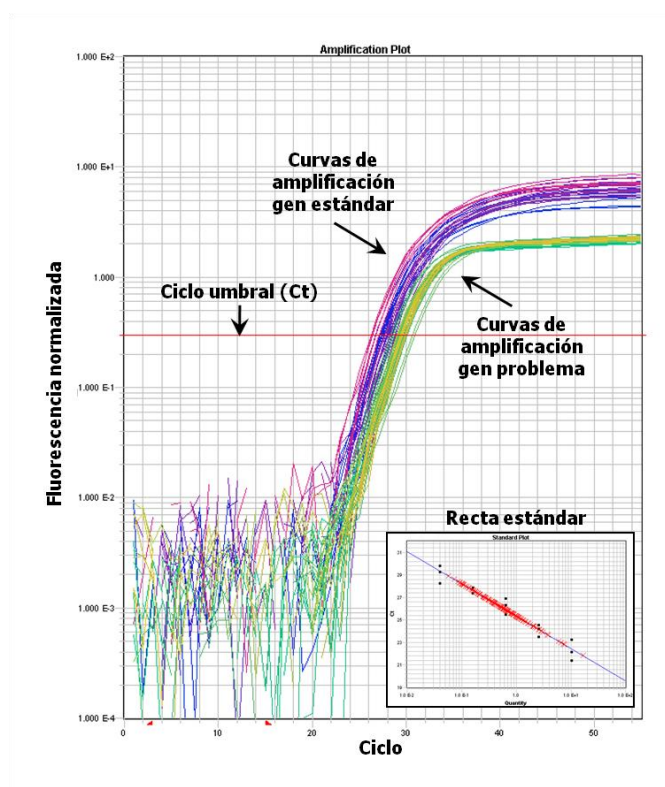
Oligonucleótido	Secuencia
p27 cebador sentido	5'-CCCTAGAGGGCAAGTACGAGT-3'
p27 cebador antisentido	5'-AGTAGAACTCGGGCAAGCTG-3'
p27 sonda UPL #39	5'-FAM-AGGTGGAG-Q-3'
GUS cebador sentido	5'-GAAAATATGTGGTTGGAGAGCTCATT-3'
GUS cebador antisentido	5'-CCGAGTGAAGATCCCCCTTTTA-3'
GUS sonda	5'-FAM-CCAGCACTCTCGTCGGTGACTGTTCA-3BQ1-3'

Cada reacción se realizó sobre un volumen total de 12 μ l, añadiendo a los 2 μ l de ADNc, 10 μ l de la siguiente mezcla, calculada por muestra:

<u>CDKN1B (p27)</u>		<u>GUS</u>	
Cebadores sent.+antisent. (5 μ M)	0.48 μ l	Cebador sentido (10 μ M)	1.20 μ l
Sonda UPL #39 (10 μ M)	0.12 μ l	Cebador antisentido (10 μ M)	1.20 μ l
Universal Master Mix	6.00 μ l	Sonda(10 μ M)	1.20 μ l
H2O	3.40 μ l	Universal Master Mix	6.00 μ l
		H2O	0.40 μ l

La PCR cuantitativa se realizó en un equipo *ABI PRISM® 7900 HT Sequence Detector* (Applied Biosystems, Foster City, CA, USA), con las siguientes condiciones de amplificación: un paso inicial de desnaturalización a 95 °C durante 10 min, seguido de 40 ciclos de 15 s a 95 °C y 1 min a 60 °C. El análisis se realizó con el programa SDS, versión 2.2.2, como se muestra en la figura M12. Se estableció un “ciclo umbral” (Ct, *cycle threshold*) en la fase inicial de la amplificación exponencial de todas las muestras, y se obtuvieron dichos valores de Ct para todas ellas. Se realizó una curva estándar, empleando diluciones seriadas 1:4 de la muestra que mostró altos niveles de ARNm de *CDKN1B*, calculándose así la expresión relativa de cada muestra con respecto a este valor. Los valores de expresión (registrados en triplicado para cada muestra) se calcularon empleando el método delta-delta Ct (Livak and Schmittgen, 2001).

Figura M12. Representación esquemática de los resultados de PCR cuantitativa



5.3. Ensayos con proteína

Para la caracterización funcional del efecto de la variante rs34330 sobre la expresión del gen *CDKN1B*, además de medir los niveles de ARNm, intentamos correlacionarlos con la expresión de proteína. Para ello, seleccionamos de forma aleatoria nueve voluntarios sanos, que representasen equilibradamente cada uno de los genotipos del SNP (CC, CT, TT), y los empleamos para realizar ensayos de expresión a nivel de proteína.

5.3.1. Extracción de proteínas

Para la extracción de proteína total de linfocitos, después de procesar las muestras como se describe en el apartado 5.1, se realizaron los siguientes pasos:

1. Descongelar el pellet de linfocitos, y disgregarlo todo lo posible mediante pulsos de vortex.
2. Preparar el tampón de lisis, que consiste en el tampón RIPA (Invitrogen, CA, USA), específico para la extracción de proteínas, al que se le añaden inhibidores de proteasas *Complete* (Roche) 1X y el detergente Igepal 1%(Sigma-Aldrich).
3. Añadir 30 µl de tampón de lisis por muestra, y dejar en agitación en una noria a 4 °C durante 1 h.
4. Centrifugar a 13000 rpm a 4 °C durante 15 min, recoger y alicuotar el sobrenadante donde están las proteínas. Congelar inmediatamente a -80 °C o mantener en hielo si se va a continuar trabajando.

5.3.2. Cuantificación de proteínas mediante Bradford

Las proteínas extraídas se cuantificaron mediante el método de Bradford, que es una técnica espectrofotométrica basada en la capacidad del colorante azul de Coomassie para unirse al enlace peptídico de las proteínas y virar de color. La técnica se basa, de forma análoga a lo descrito para el Picogreen en el apartado 2.4, en la comparación de muestras problema con diluciones seriadas de proteínas de concentración conocida.

En nuestro caso, empleamos diluciones de concentración conocida de la proteína seroalbúmina bovina (BSA), cubriendo un rango de 0 a 2 mg/ml. Las reacciones se realizaron por duplicado en placas BD Falcon™ (BD Bioscience), utilizando el reactivo *Bio-Rad Protein Assay* (Bio-Rad) y un lector *Wallac 1420* (Perkin Elmer) para medir la absorbancia a una longitud de onda de 595 nm.

5.3.3. Inmunoblot

El inmunoblot o *western blot*, es una técnica que permite la detección específica de proteínas mediante el empleo de, en primer lugar, una separación de las mismas en función de su tamaño, seguido del uso de un anticuerpo para el reconocimiento de la proteína de interés. En nuestro caso, empleamos esta técnica para la detección y cuantificación de los niveles de la proteína p27, para luego correlacionarlos con el genotipo de la variante rs34330, localizada en el promotor de *CDKN1B*, gen que codifica p27.

El procedimiento para realizar el inmunoblot comenzó con la preparación de los extractos proteicos de los nueve voluntarios sanos. Se calculó el volumen necesario para disponer de 30 µg de proteína por muestra, que se mezclaron con tampón de carga y un agente reductor (*XT Reducing Agent*, Bio-Rad), para después hervirlas durante 10 minutos, asegurando así su desnaturalización. De esta forma, las proteínas adquieren una estructura y carga similar, con lo que sólo se diferencian en su tamaño o peso molecular.

A continuación, las proteínas se separaron mediante una electroforesis vertical en un gel de poliacrilamida desnaturalizante (SDS-PAGE), a una concentración del 12 %, empleando para ello el sistema de geles y cubetas *XCell SureLock™ Mini-Cell electrophoresis system* (Invitrogen, CA, USA). La electroforesis se llevó a cabo a un voltaje constante de 200 V durante 45 min, utilizando el tampón MES. Se empleó un marcador de peso molecular Rainbow™ (GE Healthcare) para asegurar la correcta discriminación en torno a los 27 kiloDaltons (kDa), que es el peso de la proteína p27.

Terminada la electroforesis, se procedió a la transferencia de las proteínas a una membrana de polifluoruro de vinilideno (Immobilon-P Membrane, Millipore, Billerica, MA, USA), previamente activada con metanol. La transferencia se realizó en ambiente semi-seco mediante el sistema *Trans-Blot® Semi-Dry Transfer Cell* (Bio-Rad), a una intensidad de corriente constante de 300 miliamperios (mA) durante 50 min. Se comprobó la carga homogénea de las proteínas mediante la tinción de Ponceau S sobre la membrana.

Una vez separadas y transferidas las proteínas, se procedió a la detección específica de p27, mediante las siguientes incubaciones sucesivas de la membrana:

1. Un bloqueo en leche, diluida al 2.5 % en tampón TBS, durante 30 min, para evitar las posibles uniones inespecíficas posteriores.
2. Tres lavados de 5 min en tampón TBS en el que se diluye al 0.2% el detergente Tween 20.
3. Una incubación durante toda la noche a 4 °C con el anticuerpo primario anti-p27. Utilizamos una dilución 1:500 del anticuerpo sc-776 (Santa Cruz, CA, USA), que es un anticuerpo policlonal producido en conejo, que reconoce específicamente la proteína p27 íntegra (aminoácidos 1-197).
4. Tres lavados de 5 minutos con TBS-Tween 0.2 %.
5. Una incubación en oscuridad con un anticuerpo secundario fluorescente, capaz de reconocer la región constante de las inmunoglobulinas de conejo. Empleamos una dilución 1:10000 del anticuerpo *Alexa Fluor® 680 anti-rabbit IgG* (Molecular Probes, Eugene, OR, USA)
6. Tres lavados de 5 minutos con TBS-Tween 0.2 %.

Finalizadas las incubaciones, la presencia de bandas de 27 kDa se determinó escaneando la membrana en un aparato *LI-COR Odyssey® Blot scanner*, midiendo la fluorescencia a una longitud de onda de 700 nm. Para corregir la cantidad de proteína p27 en función de la cantidad de proteína total de cada muestra, se incubó la membrana con una dilución 1:10000 del anticuerpo (DM1A) (abcam, Cambridge, UK) anti- α -tubulina (α TUB), proteína de expresión ubicua empleada como control de carga.

La cuantificación de la cantidad de proteína se realizó mediante el programa Odyssey, versión 2.1 (*LI-COR*). Se midió la intensidad de las bandas de p27 y α TUB, y se corrigieron los valores sustrayendo la intensidad de sus respectivos fondos. La expresión final de p27 se expresó como unidades relativas (UR), resultado de los cocientes de intensidad p27/ α TUB.

6. Ensayos funcionales con líneas celulares

Además de los estudios con linfocitos de voluntarios sanos, realizamos ensayos funcionales más específicos con líneas celulares. Estos experimentos se emplearon para dilucidar el papel molecular de dos variantes, el rs34330 en el gen *CDKN1B* (p27), y el rs1867277 en *FOXE1*, ambos asociados de forma estadísticamente significativa al desarrollo de cáncer de tiroides. Dado que estas variantes se localizaban en región promotora, los ensayos estuvieron destinados a entender su efecto en la regulación de la expresión de sus respectivos *loci*. Para ello, se realizaron, según el caso, estudios de luciferasa y ensayos de retardo (EMSA).

6.1. Cultivos celulares

Para los estudios de luciferasa para ambas variantes, se cultivaron células HeLa (*American Type Culture Collection*), mediante incubación a 37 °C en una atmósfera húmeda con un 5 % de CO₂ y un 95 % de O₂. Utilizamos el medio de cultivo rico en glucosa *Dulbecco's Modified Eagle's Medium* (Sigma Aldrich), suplementado con 100 U/ml de penicilina, 0.1 mg/ml de estreptomina y un 10 % de suero bovino fetal.

Además, se cultivaron otras dos líneas celulares derivadas de células foliculares de tiroides: la línea PCCI3, derivada de rata, y la línea WRO, derivada de un carcinoma folicular de tiroides humano. Para el cultivo de la línea PCCI3, se utilizó el medio *F-12 Coon's medium* (Euroclone), suplementado con 100 U/ml de penicilina, 0.1 mg/ml de estreptomina, 5% de suero de ternero recién nacido ("*newborn calf serum*"), 20 ng/ml de glicil-histidil-lisina (GHL) (Sigma), 3.6 ng/ml de hidrocortisona (Sigma), 10 µg/ml de insulina de páncreas bovino (Sigma), 5 µg/ml de apo-transferrina humana (Sigma), 10 ng/ml de somatostatina (Sigma) y 0.5 mU/ml de TSH (Sigma). La línea celular WRO, se cultivó a 37 °C en una atmósfera húmeda con un 5 % de CO₂, siguiendo los métodos descritos por Schweppe y colaboradores (Schweppe, et al., 2008).

6.2. Ensayos de luciferasa

Los ensayos de actividad luciferasa nos permitieron estudiar en un sistema aislado, con ADN de cadena sencilla, el efecto de cada alelo de los SNPs de interés sobre la transcripción del gen. Además, pudimos observar la importancia de esas variantes para el efecto de otros factores.

6.2.1. Construcción de vectores reporteros

6.2.1.1. Construcción del vector pGL3-p27^{kip1}_{C/T}

El SNP rs34330 (c.-79C>T) está situado en la región promotora/5'UTR del gen *CDKN1B*. Por este motivo, se clonó la mínima región reguladora del gen *CDKN1B* que se ha demostrado que posee la capacidad de controlar eficientemente la transcripción de un gen reportero insertado a continuación (*downstream*) (Ito, et al., 1999; Liu, et al., 2005; Minami, et al., 1997). Se generaron fragmentos de ADN con la secuencia de nucleótidos comprendida entre -575 y -1 del *CDKN1B*, mediante amplificación por PCR, utilizando como molde el ADN genómico de individuos homocigotos CC o TT para la variante rs34330. Para ello, se empleó una pareja de cebadores con secuencias -79S: 5'-GGTACCGGTACCCACCTTAAGGCCGCGCTCG-3' y -79AS: 5'AAGCTTAAGCTTCTTTCTCCCGGTCTGCACG-3', que contienen dianas para las enzimas de restricción *KpnI* y *HindIII*, respectivamente. Los productos de PCR se clonaron en el sistema *pGL3-Basic promoterless plasmid* (Promega), generando así las construcciones pGL3-p27^{kip1}_C (-79C) y pGL3-p27^{kip1}_T (-79T), capaces de controlar la expresión del gen de la luciferasa (reportero), insertado *downstream*. Los plásmidos generados para cada uno de los dos genotipos se comprobaron mediante secuenciación directa.

6.2.1.2. Construcción del vector pGI3b-FOXEl-283G/A

La construcción reportera pGI3b-FOXEl incluía la región reguladora en 5' *upstream*, desde -1934 hasta +539 bp respecto al sitio de inicio de la transcripción de *FOXEl* (Eichberger, et al., 2004), y llevaba el alelo A para el SNP rs1867277 (pGI3b-FOXEl-283A). En este plásmido, el alelo A se reemplazó por el alelo G (pGI3b-FOXEl-283G) mediante mutagénesis dirigida (QuikChange II XL kit, Stratagene), siguiendo las instrucciones del fabricante y un oligonucleótido con la secuencia 5'-cagtcccggtcGcgaggccaccgc-3'. La correcta sustitución en c.-283A>G y la ausencia de artefactos se confirmó mediante secuenciación directa.

6.2.2. Transfecciones y ensayos de luciferasa

Para los ensayos referidos al gen *CDKN1B* (p27), se transfectaron las células (HeLa y PCCl3) con 1 µg de la construcción reportera promotor-luciferasa (pGL3-p27^{kip1}_C o pGL3-p27^{kip1}_T) y 100 ng de el vector control de Renilla pRL (Promega) para corregir las diferencias en la eficiencia de la transfección. Las transfecciones se realizaron utilizando el *Trans IT-LT1 Transfection Reagent* (Mirus) y las condiciones del fabricante. Para los ensayos de actividad luciferasa, las células se cosecharon a las 24 h y 48 h después de la transfección mediante centrifugación, se lavaron dos veces con PBS y se lisaron con 100 µl de agente de lisis (*Dual-Luciferase Reporter Assay System*, Promega). Los lisados celulares se dispensaron a placas luminométricas, y a continuación se realizó una inyección secuencial del compuesto *Luciferase Assay Reagent* para la medida de la actividad de la luciferasa de luciérnaga, y del agente *Stop&Glo Reagent*, para medir la actividad Renilla, respectivamente. La actividad basal de la luciferasa se midió en células transfectadas con el vector vacío (*pGL3 Basic Vector*). Se realizaron tres transfecciones independientes para cada construcción (*pGL3 Basic Vector*, pGL3-p27^{kip1}_C o pGL3-p27^{kip1}_T), y cada una de ellas, a su vez, en triplicado.

Para los ensayos de actividad del promotor de *FOXE1*, las células HeLa se transfectaron transitoriamente utilizando el sistema JetPei (PolyPlus Transfection) con distintas cantidades de vector de expresión (como se indica en cada experimento). Se emplearon 3 µg de construcción reportera y 60 ng de pRL-CMV. Cuarenta y ocho horas después de la transfección, se cosecharon las células, se lisaron y se midió la actividad de la luciferasa y la renilla. La actividad del promotor de las células transfectadas con el vector de expresión se calculó como el cociente de las actividades luciferasa:renilla relativo al cociente de las células transfectadas con el vector vacío correspondiente. Los resultados se expresaron como la media±DE de cuatro experimentos independientes, cada uno de ellos realizados en triplicado.

Además, para el estudio de *FOXE1*, se realizaron transfecciones de factores de transcripción específicos, mediante el empleo de los siguientes vectores de expresión: pcDNA3.1-DREAM (Carrion, et al., 1999), pSG5-αCREM (Foulkes, et al., 1991), pGal4-CREB (Mendez-Pertuz, et al., 2003), pN3 (USF1), pN4 (USF2) (Coulson, et al., 2003) y pUSF-1, expresando este último una forma dominante negativa de USF1. Para evaluar la eficiencia de la transfección, se empleó también el vector pRL-CMV (renilla).

6.3. Ensayos de retardo (EMSA)

6.3.1. Principios metodológicos de los ensayos de retardo

Los ensayos de retardo o EMSAs permiten definir posibles interacciones entre proteínas específicas y secuencias de ADN. Básicamente, constan de tres fases:

1. Una incubación inicial, en la que se incorporan extractos proteicos y oligonucleótidos de cadena doble marcados con γ^{32} -ATP (radiactivo). Esta incubación se realiza en unas condiciones salinas bien definidas que promueven el establecimiento de interacciones similares a las que ocurren *in vivo*.
2. Una electroforesis no desnaturalizante en gel de poliacrilamida, en la que se incluyen las muestras anteriores, así como controles negativos (ADN sin extracto proteico). En caso de producirse unión entre alguna de las proteínas del extracto proteico y la secuencia de ADN, se producirá un retraso en la migración del complejo con respecto a la cadena libre de ADN.
3. Este retraso se visualiza gracias a la presencia de radiactividad en la cadena de ADN. Para ello, se seca el gel y se coloca junto a una película autoradiográfica.

Para definir qué proteína se une de forma específica al ADN, se pueden completar estos ensayos mediante: a) la incorporación a la incubación inicial de anticuerpos frente a posibles proteínas de unión, lo que produce un mayor retardo del complejo; b) el diseño de oligonucleótidos que contengan secuencias consenso de unión de dichos factores, y que actúan como controles positivos; y c) el uso de extractos purificados de proteínas candidatas, con lo que se demuestra si realmente producen retraso al unirse al ADN (figura M13).

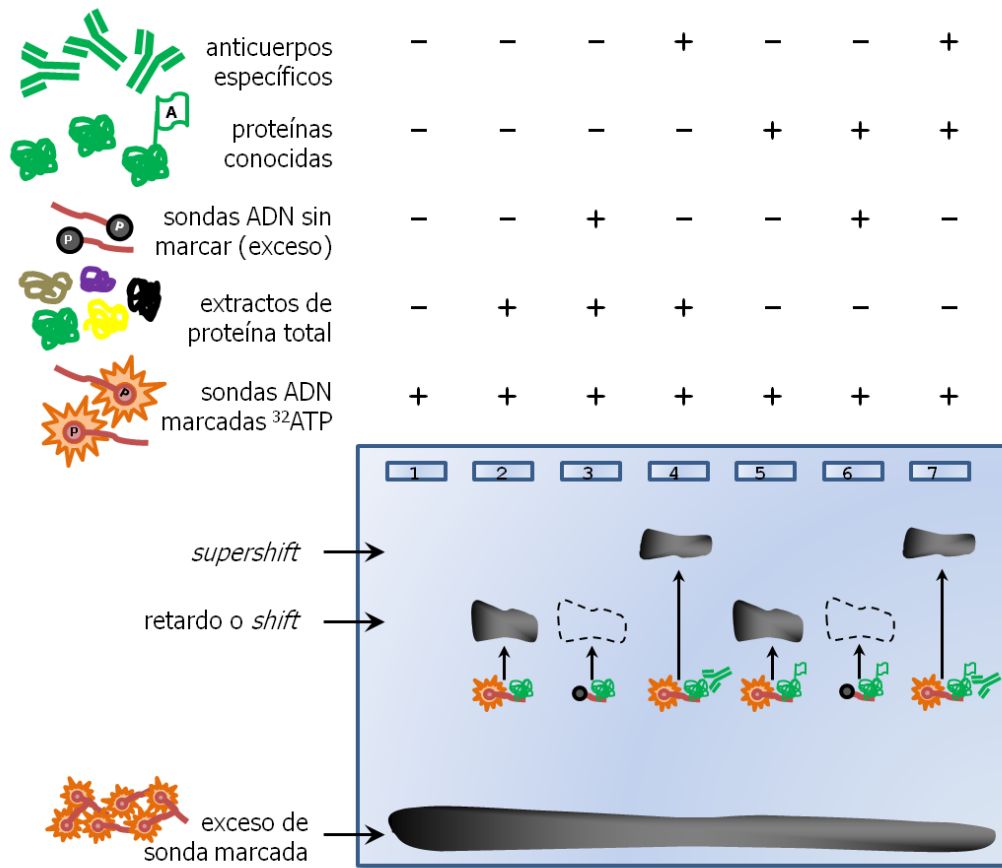


Figura M13. Esquema de los ensayos de retardo o EMSAs. La unión de proteínas a sondas de ADN, marcadas con fósforo radiactivo, y de secuencias determinadas, determina la movilidad electroforética, produciéndose un retardo cuando se produce la interacción (calles 2 y 5). La especificidad del retardo puede comprobarse mediante el empleo de un exceso de sondas no marcadas con secuencias similares o consenso (calles 3 y 6), el empleo de extractos de proteína conocida (calles 5-7) y la adición de anticuerpos específicos (calles 4 y 7).

6.3.2. Ensayos de retardo referidos a *FOXE1*

Diseñamos un EMSA para evaluar el efecto del SNP rs1867277 del promotor de *FOXE1* sobre la unión de factores de transcripción específicos. Para la obtención de los extractos proteicos se siguieron los procedimientos estándar (Andrews and Faller, 1991). Para la síntesis de proteínas específicas, se empleó el sistema TNT acoplado a lisado de reticulocitos (Promega), que permite la transcripción y traducción *in vitro*. Utilizamos los vectores de expresión de USF1/2 pN3 y pN4. Siete µg de proteína total, o alternativamente 3 µl de proteína específica obtenida mediante el sistema TNT, fueron incubados con 200 ng de las sondas de ADN de doble cadena correspondientes, previamente marcadas con γ^{32} -ATP mediante la polinucleótido kinasa T4. Las secuencias de las sondas empleadas fueron rs1867277-A: 5'-gtcccgggtcAcgaggccaccg-3' (abreviada como "alelo A"); rs1867277-G: 5'- gtcccgggtcGcgaggccaccg-3' ("alelo G"); y el elemento DRE del gen de la prodinorfina: 5'- gaagccggagtcaaggaggccctg-3' ("DRE-pDyn"). Para los ensayos de competición, se utilizó un exceso de 100 veces del mismo oligonucleótido (en cada caso) o de un oligonucleótido no relacionado (5'-ccataatgcaaaaatggaaagaattaaa-3') y no marcado ("frío"). También se utilizaron las siguientes sondas de ADN: la secuencia consenso de unión de USF (5'-cctgccacgtgacccggcct-3'; "USF-consenso"); la secuencia de unión de CRE del gen de la somatostatina (5'-cctcctagcctgacgtcagagagagagt-3'; "CRE-consenso"); y la región similar a la secuencia consenso de unión a CRE presente en el promotor de *FOXE1* (5'-accagagtcgagtcccggtcacgaggcca-3'; "CRE-*FOXE1*"). Además, se emplearon anticuerpos específicos que reconocían las proteínas DREAM, USF1, o USF2 humanas (sc-9142, sc-229 y sc-862, respectivamente, de *Santa Cruz Biotechnologies*), y se incubaron junto a los extractos proteicos y los ADN de doble cadena.

7. Estudios de interacción gen-gen o epistasia

Dado que es poco probable que una variante polimórfica en un único gen determine la susceptibilidad a desarrollar una enfermedad compleja como el cáncer de tiroides, uno de nuestros objetivos fue estudiar el efecto combinado de varias alteraciones. Para ello, se llevó a cabo un estudio de interacción gen-gen o epistasia. La epistasia puede definirse como la interacción entre variantes genéticas que da como resultado un efecto fenotípico condicionado por la acción coordinada de dichas variantes, y que no resulta de la suma de los efectos individuales de cada una de ellas (Dimas and Dermitzakis, 2009). Una representación gráfica de este efecto condicional está representada en la figura M14A.

El diseño de este estudio fue similar a un estudio caso-control clásico, pero teniendo en cuenta combinaciones de genotipos de pares de SNPs, en lugar de genotipos de una sola variante. Al igual que en el estudio individual de SNPs, se empleó la serie 1 como serie de descubrimiento, y las series 2+3 como validación de las interacciones significativas obtenidas en la primera fase.

De entre los algoritmos para la detección de epistasia disponibles, elegimos el método MB-MDR ("*Model-Based Multifactor Dimensionality Reduction*") (Calle, et al., 2008), que amplía y mejora el algoritmo MDR original (Ritchie, et al., 2001). Básicamente, MB-MDR trabaja en tres fases:

1. En primer lugar, prueba todas las combinaciones posibles de SNPs ($n=768$, en nuestro caso) elegidos dos a dos, y establece combinaciones de genotipos de 3 categorías (fig M14B): riesgo (casos significativamente sobrerrepresentados), protección (controles significativamente sobrerrepresentados) y neutro (representación similar de casos y controles). Para cada pareja de SNPs, el programa realiza un test de regresión logística, que puede ser corregido por el efecto individual de cada SNP. Además, es posible realizar un ajuste por posibles factores de confusión (edad, género y origen, en nuestro estudio).
2. A continuación, el algoritmo agrupa las combinaciones de genotipos de riesgo, protección y neutros, y los recodifica en una nueva variable, con la que realiza una nueva regresión logística.
3. Finalmente, el programa calcula el estadístico de Wald, a partir del cual, mediante un test de permutaciones, puede calcularse la significación estadística o P valor.

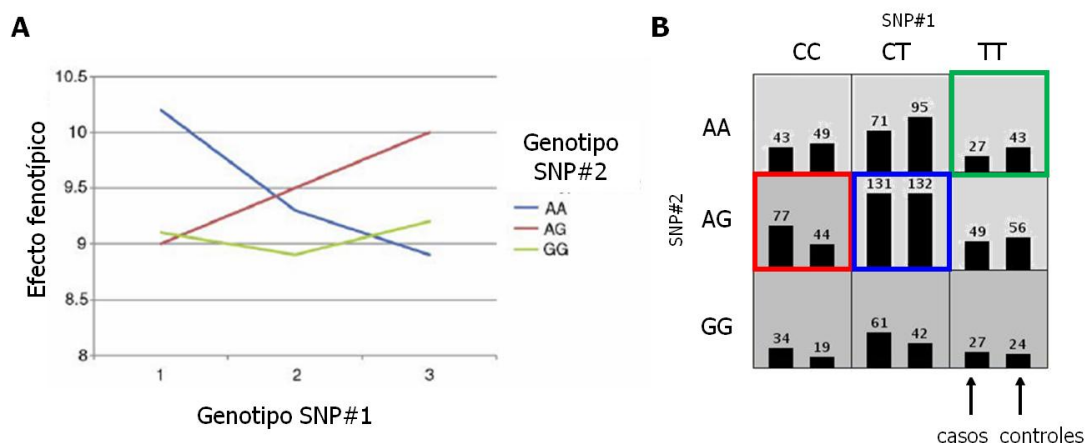


Figura M14. Análisis de interacción gen-gen o epistasia. **A.** Representación esquemática del efecto condicional sobre el fenotipo que producen dos variantes actuando de forma epistática (adaptado de Dimas & Dermitzakis, 2009); **B.** Salida del programa MDR para una pareja de SNPs, con la asignación de combinaciones de genotipos de riesgo (en rojo), protección (en verde) y neutros (en azul).

La elección de este algoritmo sobre otros se realizó al considerar que era el que más ventajas presentaba, como el hecho de no hacer ninguna selección *a priori*. La ventaja fundamental de MB-MDR, sin embargo, es que permite corregir por la contribución individual de cada SNP, de forma que el efecto observado se debe únicamente a la presencia de epistasia. Además ofrece la posibilidad de ajustar los *P* valores por posibles factores de confusión. La principal desventaja del método es que sólo considera parejas de SNPs, aunque la posibilidad de trabajar con combinaciones de más variantes estaría condicionada de todas formas por el tamaño muestral disponible, del que los estudios de epistasia, en general, son muy dependientes.

8. Análisis estadísticos

Todos los análisis estadísticos de este trabajo, salvo cuando se indica lo contrario, se realizaron empleando los programas SPSS para Windows, versión 17.0, y GraphStat.

8.1. Estudio individual de SNPs

Para el estudio de la distribución de frecuencias genotípicas entre casos y controles, primero se comprobó para cada variante que los controles estuviesen en equilibrio de Hardy-Weinberg, utilizando para ello el test exacto de Fisher. El cálculo de riesgos relativos u *odds ratios* (OR) de cada SNP y cáncer de tiroides se calculó mediante el test de X^2 (chi-cuadrado) de Pearson o mediante regresión logística binaria. Se emplearon los homocigotos para el alelo más frecuente en controles como el grupo de referencia. Para cada SNP, se estimó el modelo que mejor se ajustaba a lo observado, entre los modelos dominante, recesivo y codominante, calculándose también sus *P*-valores asociados. Los *P*-valores significativos se ajustaron por posibles elementos de confusión, como el género, la edad y el origen, mediante una regresión logística condicional.

Para evaluar la especificidad de las asociaciones estadísticas en los distintos subtipos de cáncer de tiroides, se empleó un test de cociente de probabilidades (*likelihood-ratio test*), utilizando el programa STATA[®].

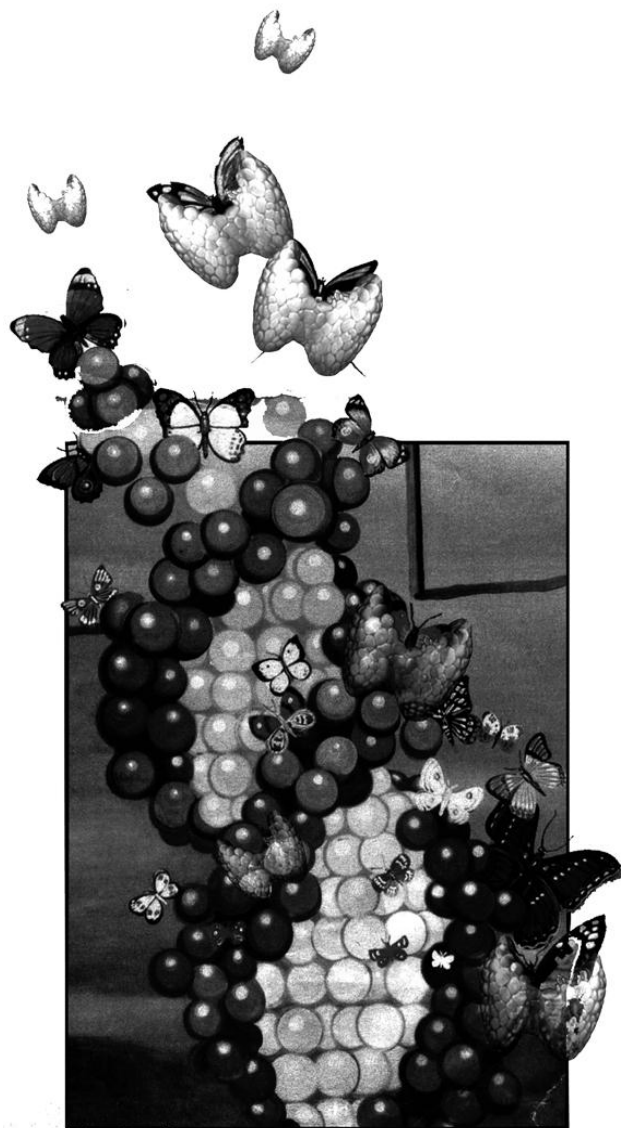
8.2. Estudio de haplotipos

Además del estudio individual de SNPs, realizamos un estudio de haplotipos o combinaciones específicas de alelos en un mismo cromosoma. Para ello, utilizamos PHASE 2.0 (Stephens, et al., 2001), una herramienta computacional basada en métodos Bayesianos. Las comparaciones caso-control de las distribuciones de haplotipos se realizaron aplicando una corrección mediante un test de 10000 permutaciones. Los tests de asociación entre haplotipos específicos y riesgo a desarrollar cáncer de tiroides se realizaron empleando el paquete Haplo.Stats en R (<http://www.R-project.org>).

8.3. Análisis estadísticos de los estudios funcionales

Los resultados obtenidos en los ensayos de luciferasa se analizaron mediante comparación de medias empleando el test de t de student de dos colas. Los resultados de PCR cuantitativa, al no cumplir los requisitos de normalidad, se analizaron por comparación de medianas mediante un test de U de Mann-Whitney.

RESULTADOS



La sección de resultados está organizada cronológicamente, y en función de la disponibilidad de muestras en cada momento.

En el primer apartado, se muestran los resultados obtenidos para el gen *CDKN1B* (p27), referidos al conjunto de casos y controles españoles disponibles en el momento del estudio (serie 1 incompleta). Además, se realizó una caracterización funcional de la variante asociada.

El segundo apartado muestra los resultados de asociación del estudio de 97 genes candidatos en la fase de descubrimiento, o serie 1 de casos y controles.

La tercera sección está dedicada exclusivamente a los resultados obtenidos para el gen *FOXE1*. Se ha decidido separar estos resultados del apartado anterior dado su extensión e importancia. Se incluyen los resultados de asociación de las series 1 y 2, junto a la caracterización funcional de la variante asociada.

En el apartado cuarto, se muestran los resultados de validación en las series 2 y 3 de las 11 variantes candidatas obtenidas en el apartado 2.

En la quinta y última sección, se muestran los resultados del estudio de interacción gen-gen o epistasia.

1. Resultados referidos al gen CDKN1B (p27^{kip1})

1.1. El polimorfismo rs34330 (-79C>T) está significativamente asociado con la variante folicular del carcinoma papilar de tiroides (CPTvf)

Las variantes -838C>A (rs36228499), -79C>T (rs34330) y 326T>G (p.V109G, rs2066827) se genotiparon en los casos y controles españoles disponibles en ese momento, para determinar si existía alguna asociación significativa entre ellos y el cáncer de tiroides. Los controles cumplieron el equilibrio de Hardy-Weinberg para los tres SNPs, como puede observarse en la Tabla R1.

Tabla R1. Polimorfismos de *CDKN1B* y muestras genotipadas

SNP (cambio nt)	Localización	Alelos (MAF)	Controles (HWE ^b)	CPT	CPTc	CPTvf	CFT
rs2066827 (T326G)	Exón 1 V109G	T/G (0.20)	334 (0.27)	313	181	100	25
rs34330 (-79 C>T)	5' UTR / Promotor	C/T (0.21)	385 (1.00)	328	178	129	47
rs3622849 (-838 C>A)	Promotor	C/A (0.46) ^c	296 (0.64)	326	193	108	23

^a Las muestras sin genotipo fueron excluidas; ^b *P* valor asociado al equilibrio de Hardy-Weinberg en controles; ^c Frecuencia en nuestra serie de controles (no había frecuencias descritas para este SNP en bases de datos públicas); Abreviaturas: nt=nucleótido; MAF=frecuencia del alelo menor; CPT=carcinoma papilar de tiroides; CPTc=CPT clásico; CPTvf=variante folicular del CPT; CFT=carcinoma folicular de tiroides; UTR= región no traducida (*untranslated region*).

Mientras que las variantes p.V109G y -838C>A no mostraron asociación significativa para ninguno de los subtipos de cáncer de tiroides analizados, identificamos el SNP con rs34330 (-79C>T) como un factor de riesgo para el desarrollo de la variante folicular del carcinoma papilar de tiroides (CPTvf). De acuerdo con estos resultados, el genotipo polimórfico TT del rs34330 está significativamente sobrerrepresentado en los pacientes con CPTvf. La variante cumplió un modelo recesivo (OR=2.12; IC 95%=1.09-4.15; *P*=0.023; Tablas R2 y R3).

Esta asociación permaneció estadísticamente significativa después de corregir nuestros resultados en función del género y la edad de casos y controles. Primero comprobamos que el OR se mantuvo homogéneo (*P*>0.05) mediante el test de Tarone. Luego, demostramos que ni el género ni la edad estaban actuando como factores de confusión de nuestros resultados, primero mediante el test de Mantel-Haenszel, y

también mediante una regresión logística condicional por pasos, con la que se obtuvo un OR ajustado=2.01 y $P=0.041$. Por último, comprobamos que el rs34330 se asociaba específicamente al subtipo CPTvf y no al resto de CPT, mediante un *likelihood-ratio test* (LR $P=0.0009$).

Debido a que en el momento de este estudio no disponíamos de grupos de validación de nuestros resultados (series 2 y 3), no fue posible intentar replicar la asociación. Sin embargo, realizamos una caracterización funcional, como se explica en los siguientes apartados, para tratar de explicar el mecanismo por el cual esta variante conferiría un riesgo a desarrollar CPTvf.

Tabla R2. Riesgos a desarrollar cáncer de tiroides en función del genotipo de rs34330 (-79C>T)

Grupo	N	Modelo codominante				Modelo recesivo		p ^b
		CC <i>vs.</i> CT		CC <i>vs.</i> TT		(CC+CT) <i>vs.</i> TT		
		OR ^a	IC 95%	OR ^a	IC 95%	OR ^a	IC 95%	
Controles	385							
CPT	328	0.96	0.70-1.32	1.44	0.81-2.56	1.46	0.83-2.56	0.186
CPTc	178	0.91	0.63-1.33	1.05	0.51-2.18	1.09	0.53-2.23	0.819
CPTvf	129	1.08	0.70-1.65	2.20	1.10-4.38	2.12	1.09-4.15	0.023
CFT	47	1.39	0.75-2.59	0.78	0.17-3.49	0.67	0.15-2.92	0.590

^a OR obtenidos al comparar casos frente a controles considerando el alelo T como el alelo de riesgo; ^b Los P valores son los asociados al modelo recesivo; Abreviaturas: N=número de muestras; OR=*odds ratio*; IC 95%=intervalo de confianza al 95 %; CPT=carcinoma papilar de tiroides; CPTc=CPT clásico; CPTvf=variante folicular del CPT; CFT=carcinoma folicular de tiroides. Los resultados estadísticamente significativos se muestran en negrita.

Tabla R3. Riesgos estimados para los SNPs de *CDKN1B* en la variante folicular del carcinoma papilar de tiroides (CPTvf)

SNP (localización)	Alelo P (MAF)	CPTvf (N)	Modelo codominante				Modelo recesivo		<i>p</i> ^b
			SS <i>vs.</i> PP		SS <i>vs.</i> PP		(SS+SP) <i>vs.</i> PP		
			OR ^a	IC 95%	OR ^a	IC 95%	OR ^a	IC 95%	
rs36228499 (-838 C>A)	A (0.46)	108	1.00	0.55-1.79	1.16	0.61-2.19	1.16	0.72-1.87	0.544
rs34330 (-79 C>T)	T (0.21)	129	1.08	0.70-1.65	2.20	1.10-4.38	2.12	1.09-4.15	0.023
rs2066827 (T326G)	G (0.20)	100	0.80	0.50-1.28	1.24	0.52-2.98	1.37	0.58-3.20	0.474

^a OR obtenidos al comparar casos frente a controles considerando el alelo polimórfico (P) como el alelo de riesgo. ^b Los P valores son los asociados al modelo recesivo; Abreviaturas: S=alelo *wildtype*; P=alelo polimórfico; N=número de muestras; MAF=minor allele frequency; OR= *odds ratio*; IC 95%=intervalo de confianza al 95 %; CPTvf=variante folicular del carcinoma papilar de tiroides. Los resultados estadísticamente significativos se muestran en negrita.

1.2. La variante -79C>T disminuye la tasa de transcripción de *CDKN1B* en ensayos de luciferasa

Dado que el genotipo -79TT del promotor del gen *CDKN1B* (p27^{kip1}) estaba asociado con un aumento de riesgo estadísticamente significativo a desarrollar CPTvf, se intentó determinar si la variante -79C>T tenía alguna consecuencia funcional sobre la expresión de p27.

En una primera serie de experimentos, se determinó la capacidad transcripcional de las dos versiones posibles del promotor de *CDKN1B*, transfectando cada una de las construcciones de plásmido-promotor, es decir, pGL3-p27^{kip1}_C (-79C) y pGL3-p27^{kip1}_T (-79T), a células HeLa. Después de 48 h en presencia o ausencia de suero, se recolectaron las células y se midió la actividad transcripcional de las dos variantes de promotor mediante un ensayo de luciferasa (fig R1A). La actividad luciferasa se normalizó con el control interno de Renilla y la actividad basal del vector pGL3 vacío. Se observó que la actividad transcripcional mostrada por el promotor polimórfico (p27^{kip1}_T) era significativamente menor que la exhibida por el promotor *wildtype* (p27^{kip1}_C), tanto en presencia ($P=0.001$) como en ausencia de suero ($P=0.002$).

Posteriormente, se determinó la actividad de las diferentes variantes del promotor de p27. Para ello, se transfectaron células HeLa con las construcciones pGL3-p27^{kip1}_C y pGL3-p27^{kip1}_T, y se sometieron a condiciones de falta de nutrientes ("starvation"), utilizando un medio de cultivo libre de suero durante 24 h. A continuación, reemplazando del medio anterior por otro fresco y enriquecido con suero al 10 %, se indujo a estas células a volver a entrar en el ciclo celular, y se midió la actividad luciferasa después de 0, 12 y 24 h. Después de normalizar la actividad luciferasa con el control interno de Renilla, la actividad transcripcional media de pGL3-p27^{kip1}_T fue consistentemente menor (aproximadamente un 50 %) que la mostrada por pGL3-p27^{kip1}_C (fig R1B). La actividad del promotor *wildtype* de *CDKN1B* (pGL3-p27^{kip1}_C) aumentó hacia la mitad de la fase G1 del ciclo (12 h tras la administración de suero) y disminuyó en la transición G₁/S (24 h después de la administración de suero). Por su parte, el promotor polimórfico (pGL3-p27^{kip1}_T) mostró una cinética similar, pero con una actividad transcripcional significativamente reducida ($P_{0h}=0.002$, $P_{12h}<0.001$ y $P_{24h}<0.001$, respectivamente).

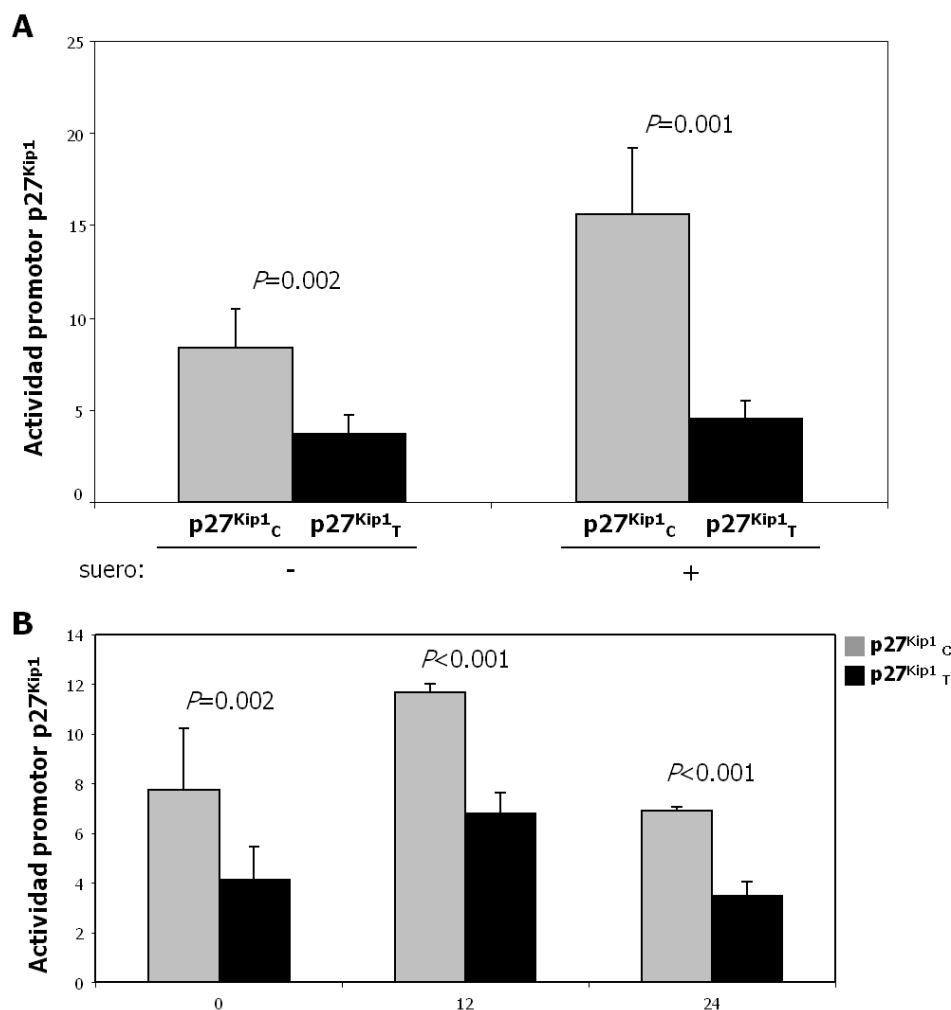


Figura R1. Actividad transcripcional de las construcciones controladas por el promotor del gen *CDKN1B*, conteniendo las variantes -79C o -79T, en células HeLa. **A.** Actividad luciferasa corregida de las construcciones pGL3-p27^{Kip1}_C y pGL3-p27^{Kip1}_T con y sin suero. Los datos representan la media±desviación típica de tres transfecciones independientes. **B.** Actividad transcripcional de pGL3-p27^{Kip1}_C y pGL3-p27^{Kip1}_T en células HeLa en función del ciclo celular. Los tiempos 0, 12 y 24 h se refieren a los momentos de recolección y medida de la actividad de luciferasa tras la re-administración de suero para la entrada en el ciclo celular.

Posteriormente, analizamos la actividad del promotor de *CDKN1B* en células foliculares de tiroides, o tirocitos. Para ello, empleamos como modelo las células PCCI3, que constituyen una línea establecida de células diferenciadas de tiroides de rata. Transfectamos estas células con las construcciones pGL3-p27^{Kip1}_C y pGL3-p27^{Kip1}_T, se cultivaron en un medio con o sin TSH durante 24 y 48 h, y se midió la actividad luciferasa. Tras la normalización de la actividad luciferasa mediante el control Renilla y

la actividad basal del vector pGL3 vacío, la actividad transcripcional de pGL3-p27^{kip1}_T en tirocitos fue aproximadamente el 50 % de la mostrada por pGL3-p27^{kip1}_C (P valores <0.001, fig R2). Curiosamente, la diferencia en la actividad transcripcional fue más evidente cuando las células habían sido privadas de TSH.

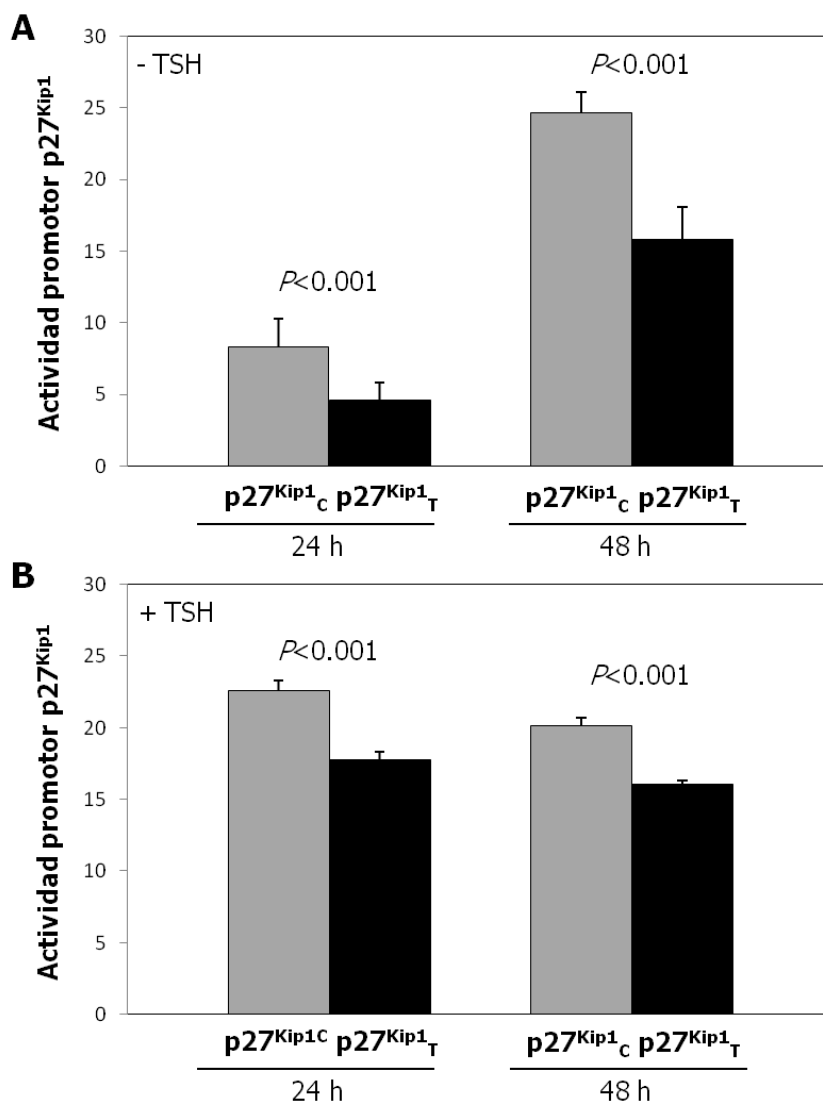


Figura R2. Actividad transcripcional de las construcciones controladas por el promotor del gen *CDKN1B*, conteniendo las variantes -79C o -79T, en células de tiroides de rata. Los tiempos 24 y 48 h son el periodo de cultivo de las células en sus correspondientes medios. Los datos representan la media \pm desviación típica de tres transfecciones independientes. **A.** Actividad luciferasa corregida de las construcciones pGL3-p27^{Kip1}_C y pGL3-p27^{Kip1}_T en un medio sin TSH. **B.** Actividad luciferasa corregida de las construcciones pGL3-p27^{Kip1}_C y pGL3-p27^{Kip1}_T en un medio suplementado con TSH.

1.3. El genotipo de riesgo -79TT se correlaciona con menores niveles de ARN mensajero de *CDKN1B* en linfocitos

Con el objetivo de confirmar los datos *in vitro* de los ensayos de transfección, evaluamos la expresión en ARN mensajero (ARNm) de *CDKN1B* mediante PCR cuantitativa a tiempo real. Para ello, utilizamos linfocitos extraídos de sangre periférica de una serie de 97 voluntarios sanos con genotipo conocido para la variante -79C>T. Las frecuencias genotípicas en este grupo fueron CC=0.59, CT=0.36 y TT=0.05, similares a las descritas para la población HapMap-CEU. Los valores de Ct medio obtenidos para *CDKN1B* y GUS fueron 29.4 y 27.7, respectivamente. Cuando comparamos la expresión de los portadores del genotipo CC frente a TT, se obtuvo una diferencia con una significación estadística marginal ($P=0.069$, fig R3A). Esta tendencia se observó también al agrupar los genotipos de no-riesgo (CC+CT) y compararlos con el grupo de riesgo TT ($P=0.102$, fig R3B). Estos resultados refuerzan la idea de que el genotipo -79TT se asocia con una reducción en la expresión de ARNm de *CDKN1B*.

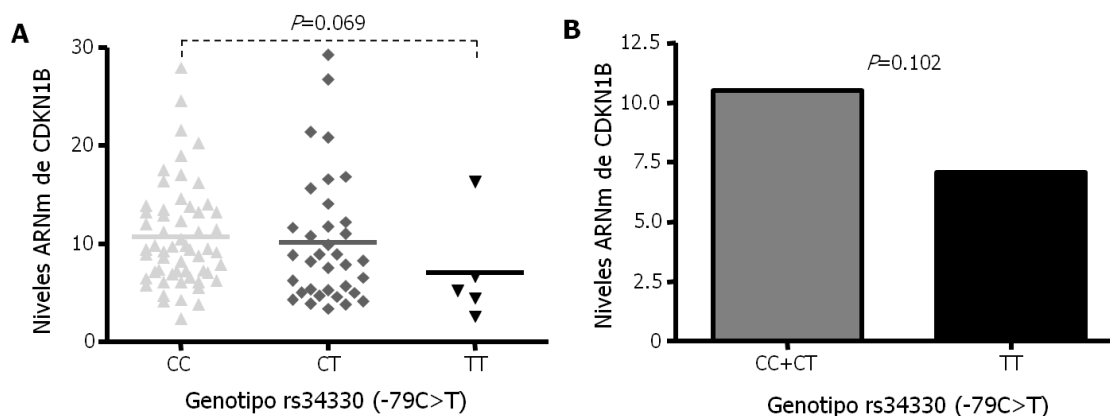


Figura R3. Expresión diferencial de *CDKN1B* en linfocitos humanos, distribuidos de acuerdo al genotipo del polimorfismo -79C>T (rs34330). **A.** Niveles de ARN mensajero (ARNm) de *CDKN1B* en linfocitos extraídos de 97 voluntarios sanos, dependiendo del genotipo de la variante promotora -79C>T. Cada punto representa la media, para cada muestra, de los niveles normalizados de ARNm, medidos en triplicado, mediante PCR cuantitativa a tiempo real. Las líneas horizontales representan el valor medio para cada grupo de genotipos. **B.** Comparación de los niveles de ARNm de *CDKN1B* entre los portadores de genotipos CC+CT frente a TT. La diferencia de expresión entre ambos grupos refuerza el modelo recesivo propuesto en el estudio de asociación.

1.4. Los portadores del genotipo -79TT muestran una tendencia hacia una menor expresión de proteína p27 en linfocitos

Para comprobar si la disminución en los niveles de de ARNm de *CDKN1B* en función del genotipo del rs34330 resultaba en niveles distinguibles de proteína, se realizaron ensayos de inmunoblot. Para ello, se emplearon extractos proteicos de nueve de los voluntarios sanos, elegidos aleatoriamente, pero que representasen de forma equilibrada los tres genotipos (CC, CT, TT). Los resultados, representados en la figura R4, sugieren el efecto del SNP en los niveles de proteína, observándose una menor expresión de p27 en los portadores del genotipo TT, aunque la magnitud de esta disminución es modesta.

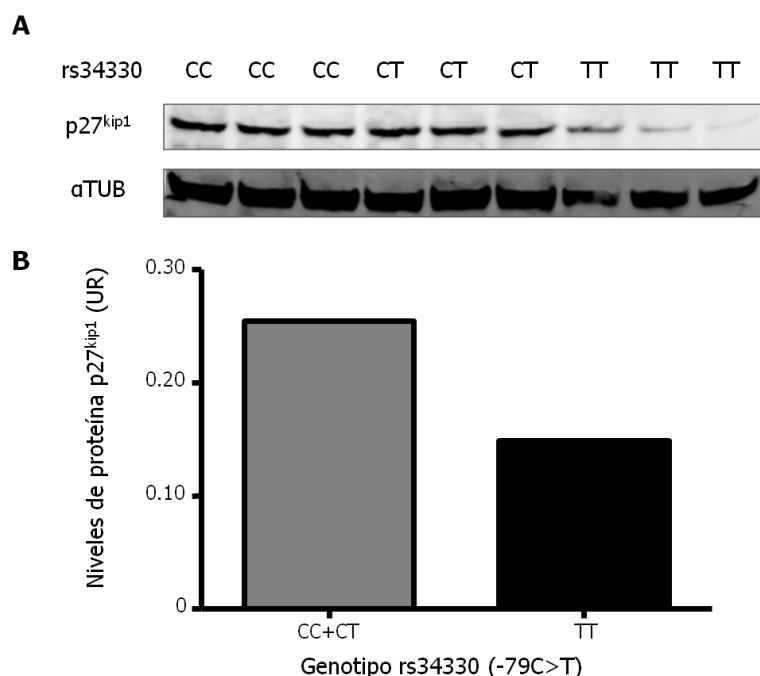


Figura R4. Niveles de proteína p27^{kip1} en linfocitos humanos, evaluados mediante inmunoblot, en función del genotipo del polimorfismo -79C>T (rs34330). **A.** Imagen del inmunoblot, mostrando la expresión de p27^{kip1} en lisados de linfocitos de voluntarios sanos, dependiendo del genotipo de la variante promotora -79C>T. La carga inicial de proteína total fue de 30 µg en todas las calles del gel, y se verificó empleando la α-tubulina (αTUB) como referencia. **B.** Comparación de los niveles de proteína p27^{kip1} entre los portadores de genotipos CC+CT frente a TT. Se observa una tendencia hacia una menor expresión de proteína en el grupo de riesgo TT. Los niveles de expresión de p27^{kip1} están expresados en unidades relativas (UR), resultado de corregir la expresión de p27^{kip1} con los niveles de α-tubulina para cada muestra (cociente p27^{kip1}/αTUB).

2. Resultados del SNP array de la primera fase del estudio de asociación

2.1. Resultados globales del genotipado en la serie 1

2.1.1. Resultados técnicos del genotipado

De la lista inicial de 768 variantes estudiadas en la serie 1, 33 SNPs (el 4 %) fueron descartados por motivos de calidad de la técnica de genotipado. Las causas de error fueron: 1. la ausencia de señal o la presencia de señales demasiado bajas; 2. la presencia de más de tres *clusters* o grupos de genotipos; y 3. el fallo en la capacidad de discriminación alélica consistente de la sonda. Esta tasa de error en el genotipado se ajustó a lo descrito para la plataforma empleada (Illumina GoldenGate®), que prevé una proporción en torno al 5 % de tasa de error. Además, otros 45 polimorfismos (un 6 %) fueron excluidos de los análisis posteriores porque sus frecuencias en el grupo control se desviaron significativamente del equilibrio de Hardy-Weinberg ($P < 0.05$). Finalmente, un total de 690 SNPs fueron considerados en el estudio (figura R5), y sobre ellos se realizaron los análisis individuales y de distribución de haplotipos pertinentes. Los códigos y posiciones de estos 690 SNPs pueden consultarse en la tabla S1. Para estos polimorfismos, la concordancia de genotipos de los duplicados inter e intraplaca fue del 100 %.

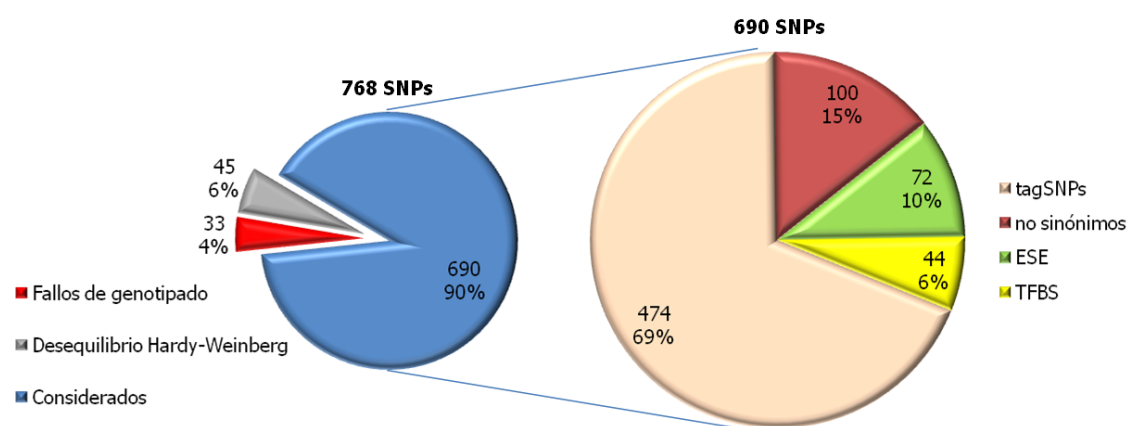


Figura R5. Representación gráfica de los resultados globales del genotipado de variantes polimórficas en la serie 1. De los 768 SNPs iniciales, 690 variantes fueron finalmente consideradas para análisis posteriores. La imagen de la derecha muestra los tipos de SNPs representados en ese grupo. Abreviaturas: ESE= SNP localizado en *Exonic Splicing Enhancers*; TFBS= SNP en *Transcription Factor Binding Sites*.

2.1.2. Resultados globales según las frecuencias alélicas

Dado que en nuestro array estaban representados polimorfismos que cubrían uniformemente todo el rango de frecuencias posible, nos preguntamos, en primer lugar, la capacidad de nuestro ensayo para detectar diferencias significativas dependiendo de la frecuencia alélica de los SNPs estudiados. Como se observa en la figura R6, si bien existe una tendencia hacia una lógica mayor capacidad para detectar diferencias para variantes con MAF elevado, parece que nuestro estudio es capaz de detectar asociaciones en todo el rango de frecuencias considerado.

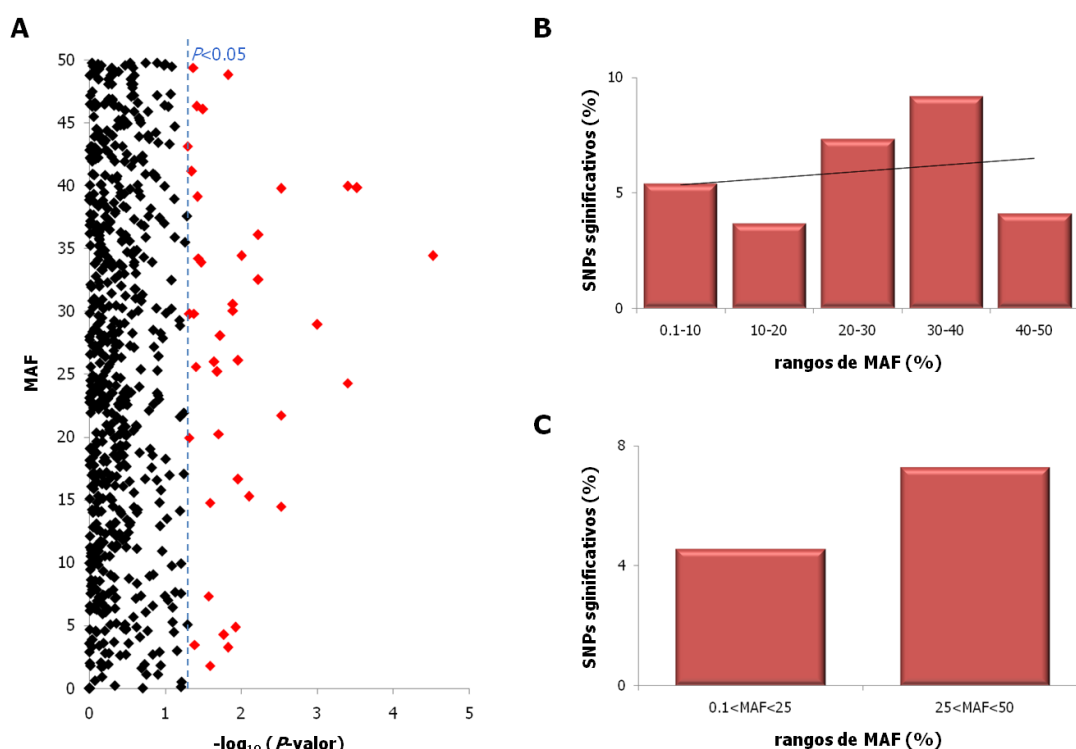


Figura R6. Resultados globales de la distribución de frecuencias alélicas entre casos ($n=609$) *versus* controles ($n=504$) en función de la frecuencia (MAF) de las variantes estudiadas. **A.** Representación gráfica de los resultados para los 690 polimorfismos considerados, en función de su MAF. En rojo, se muestran los SNPs significativos ($P < 0.05$, línea azul de puntos) para esta comparación. **B.** Porcentaje de SNPs significativos en cada rango de frecuencias alélicas (n SNPs significativos/ n SNPs en ese rango). Se muestra una línea de tendencia. **C.** Porcentaje de SNPs significativos entre SNPs poco frecuentes ($MAF < 25\%$) y más frecuentes ($MAF > 25\%$). Abreviaturas: MAF= *Minor Allele Frequency*.

2.2. Estudio individual de polimorfismos

Con los 690 polimorfismos que mostraron calidad de genotipado suficiente, realizamos una comparación de la distribución de sus frecuencias alélicas y genotípicas entre casos y controles, destinada a identificar genes de baja penetrancia (GBP) asociados al desarrollo de cáncer de tiroides.

En primer lugar, realizamos un análisis global de casos frente a controles, cuyos resultados están resumidos en la figura R7, en la que pueden verse cómo algunas variantes no se comportan conforme a lo esperado por azar. Dado nuestro interés en identificar genes específicos de cada tipo y subtipo tumorales de carcinomas de tiroides de célula folicular, a continuación se presentan los resultados de asociación en función de cada grupo de tumores.

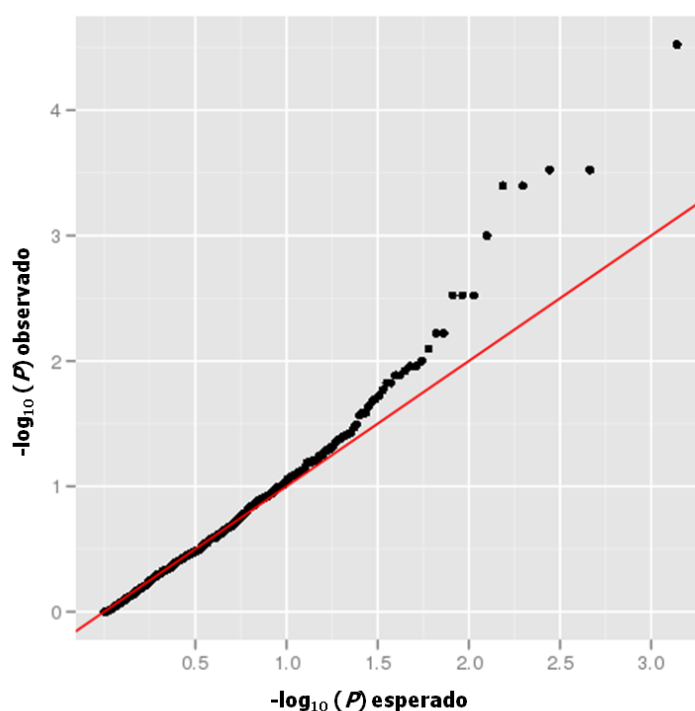


Figura R7. Gráfico Q-Q de la distribución de las frecuencias alélicas de los 690 SNPs estudiados en casos frente a controles. Se observa cómo un grupo de SNPs se desvía del comportamiento esperado por azar.

2.2.1. Resultados en Carcinoma Papilar de Tiroides (CPT)

Los resultados significativos de la comparación entre casos de carcinoma papilar de tiroides (CPT) y controles en la serie 1 están resumidos en la figura R8 y en la tabla R4. Como puede observarse, se encontraron resultados muy significativos para variantes situadas en el *locus* de *FOXE1*. Dada la importancia y extensión de los resultados obtenidos para *FOXE1*, hemos dedicado una sección completa exclusivamente a este gen (apartado 3). Además, encontramos significación estadística para polimorfismos pertenecientes a *TSHR* y *TG*, aunque, como se muestra en los siguientes apartados (2.2.2 y 2.2.3), estos genes parecen ser específicos de determinados subtipos de CPT. Entre los genes asociados al desarrollo de CPT en esta fase del estudio, destacan también por la consistencia de los resultados, *PRKAR1A* (que codifica una subunidad reguladora de la kinasa dependiente de AMPc), *SMAD7* (una proteína nuclear que inhibe la vía de TGF- β) y *KIT* (homólogo humano del proto-oncogén c-kit).

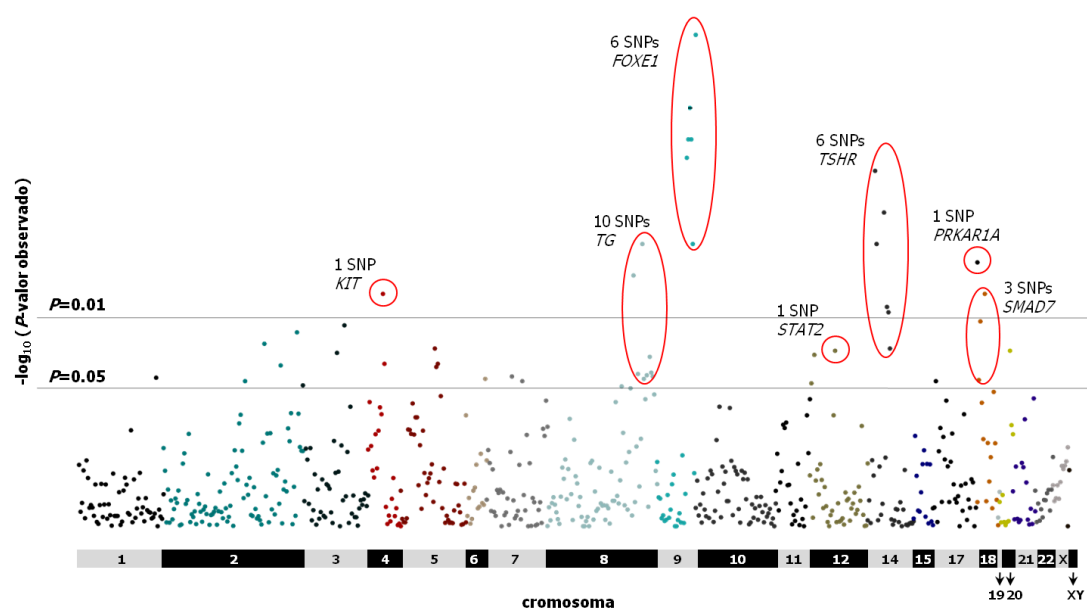


Figura R8. Manhattan plot de la distribución de frecuencias alélicas entre CPT y controles

Tabla R4. Resultados significativos ($P < 0.05$) al comparar CPT frente a controles en la serie 1.

Cr	Coord	SNP	Gen	Posición	Función	Alelos	P^a
9	99664423	rs10119760	FOXE1	downstream	tagSNP	C/G	2×10^{-5}
9	99653961	rs3758249	FOXE1	upstream	TFBS	A/G	1×10^{-4}
9	99652091	rs894673	FOXE1	upstream	tagSNP	A/T	2×10^{-4}
9	99654938	rs907577	FOXE1	upstream	TFBS	A/G	2×10^{-4}
9	99645254	rs7048394	FOXE1	upstream	tagSNP	C/T	3×10^{-4}
14	80623486	rs2284734	TSHR	intrónico	tagSNP	A/G	4×10^{-4}
14	80639431	rs1017141	TSHR	intrónico	tagSNP	A/G	1×10^{-3}
14	80623539	rs2284735	TSHR	intrónico	tagSNP	A/G	2×10^{-3}
8	134191878	rs2979033	TG	intrónico	tagSNP	A/G	2×10^{-3}
9	99656842	rs3021526	FOXE1	exónico	p.Ser275Ser	C/T	2×10^{-3}
17	64045740	rs16973034	PRKAR1A	downstream	tagSNP	A/G	3×10^{-3}
8	134052230	rs10106773	TG	intrónico	tagSNP	A/G	4×10^{-3}
18	44715010	rs17186877	SMAD7	intrónico	tagSNP	A/G	6×10^{-3}
4	55249088	rs2703488	KIT	intrónico	tagSNP	A/G	6×10^{-3}
14	80641412	rs7151769	TSHR	intrónico	tagSNP	C/T	8×10^{-3}
14	80644758	rs3783941	TSHR	exónico	p.Arg248Ser	G/T	9×10^{-3}
18	44712225	rs9946510	SMAD7	intrónico	tagSNP	A/C	0.011
3	12621255	rs6777175	RAF1	intrónico	TFBS	A/T	0.012
2	136584543	rs13022389	CXCR4	downstream	tagSNP	C/G	0.018
5	90015506	rs16868972	GPR98	exónico	p.Leu2004Phe	G/T	0.020
14	80645172	rs2080305	TSHR	intrónico	tagSNP	C/T	0.020
12	55026949	rs2066807	STAT2	exónico	p.Met594Ile	C/G	0.021
20	35455953	rs6018257	SRC	intrónico	tagSNP	C/T	0.021
3	12599763	rs7956	RAF1	downstream	tagSNP	A/G	0.022
12	6042463	rs1800378	VWF	exónico	p.His484Arg	A/G	0.023
8	134199608	rs10096431	TG	intrónico	tagSNP	A/G	0.024
5	90024260	rs2366926	GPR98	exónico	p.Asn2345Ser	A/G	0.028
4	55266198	rs3111800	KIT	intrónico	tagSNP	C/T	0.028
2	196753322	rs1378624	STK17B	upstream	tagSNP	A/G	0.029
5	90021638	rs10037067	GPR98	exónico	p.Tyr2232Cys	A/G	0.030
8	134209728	rs17705719	TG	intrónico	tagSNP	C/T	0.034
8	134062846	rs3892189	TG	intrónico	tagSNP	C/T	0.035
8	134194211	rs2979040	TG	intrónico	tagSNP	A/G	0.036
8	134212816	rs2979051	TG	intrónico	tagSNP	A/G	0.037
1	155091425	rs4661063	NTRK1	intrónico	TFBS	A/G	0.038
8	134192567	rs4736435	TG	intrónico	tagSNP	A/C	0.039
6	87870842	rs6909901	CGA	upstream	tagSNP	A/T	0.039
18	44707461	rs4939827	SMAD7	intrónico	tagSNP	C/T	0.040
7	107104411	rs10276013	SLC26A4	intrónico	tagSNP	A/G	0.041
2	113700974	rs2241975	PAX8	intrónico	tagSNP	A/G	0.041
12	6014245	rs216321	VWF	exónico	p.Arg852Gln	C/T	0.043
3	4691811	rs2306877	ITPR1	exónico	ESE	A/C	0.045
8	133991009	rs2687834	TG	intrónico	tagSNP	A/C	0.046
8	134044465	rs2069561	TG	exónico	p.Asp1838Asn	A/G	0.048

La tabla está ordenada por P -valor (de más a menos significativo); ^a P -valor asociado a la diferencia de frecuencias alélicas; Abreviaturas: Cr= cromosoma; Coord= coordenada; ESE= Exonic Splicing Enhancers; TFBS= Transcription Factor Binding Sites.

2.2.2. Resultados en Carcinoma Papilar de Tiroides clásico (CPTc)

A continuación, realizamos análisis específicos estratificando el grupo de casos en función del subtipo tumoral. Para la primera comparación, se empleó el subtipo clásico de los carcinomas papilares de tiroides (CPTc), representado en nuestra serie 1 por 305 muestras, lo que suponía casi dos tercios de los CPT.

Los resultados de asociación muy significativos para la comparación CPTc *vs.* controles están resumidos en la tabla R5. Además de *FOXE1*, que será comentado más adelante, se obtuvieron resultados altamente significativos para variantes situadas en el *locus* del gen del receptor de la tirotrópina, o *TSHR*, sugiriendo el papel de éste como un GBP en CPTc (figura R9).

La especificidad de la asociación de *TSHR* con CPTc se comprobó posteriormente al comparar los resultados obtenidos para el mejor tagSNP (rs2284734) de *TSHR* en CPTc con los de CPTvf mediante un *likelihood-ratio (LR) test* (LR $P=0.0015$).

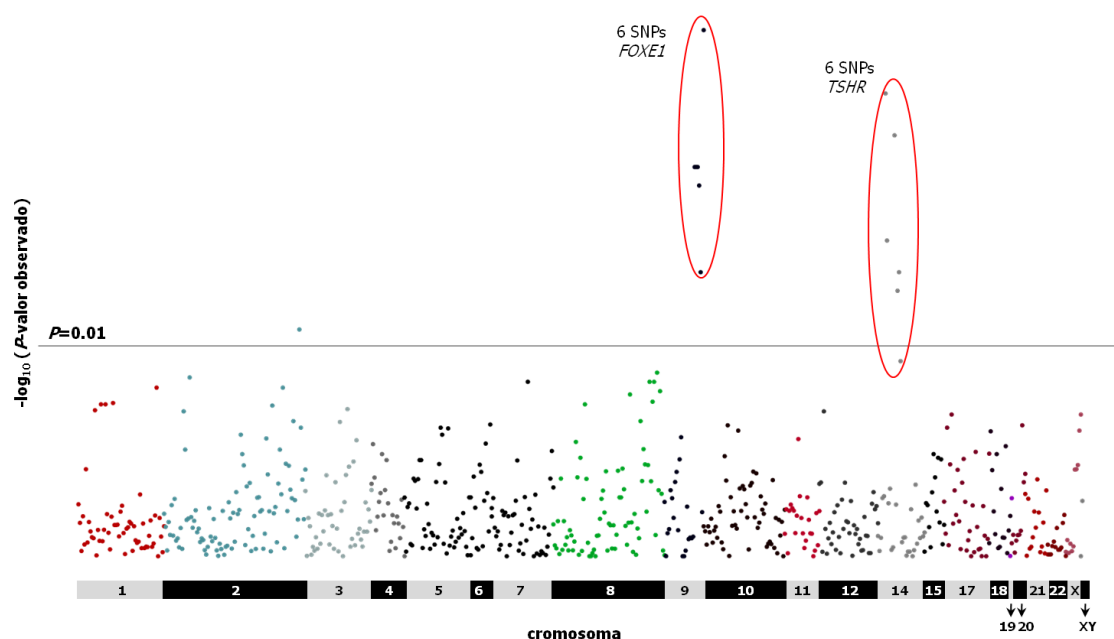


Figura R9. Manhattan plot de la distribución de frecuencias alélicas entre CPTc y controles

2.2.3. Resultados en la variante folicular del Carcinoma Papilar de Tiroides, (CPTvf)

Otro de los subtipos de CPT bien representado en nuestra serie fue la variante folicular de CPT (CPTvf), cuyas características parecen estar entre un CPT y CFT, quedando claro que es una entidad que debe ser considerada aparte. Al estratificar la muestra en CPTvf (n=146, casi un tercio de los CPT) y compararla frente a controles, obtuvimos resultados muy significativos para genes diferentes de los que aparecían en CPTc, apoyando así esta especificidad, y por lo tanto el valor de la subestratificación (figura R10). Los más significativos se obtuvieron en variantes situadas en el gen de la tiroglobulina (*TG*) que podría actuar como GBP en este subtipo (tabla R6).

De nuevo, se comprobó que la asociación para el mejor tagSNP de *TG* (rs2687834) era específica del subtipo CPTvf (LR $P=0.0023$).

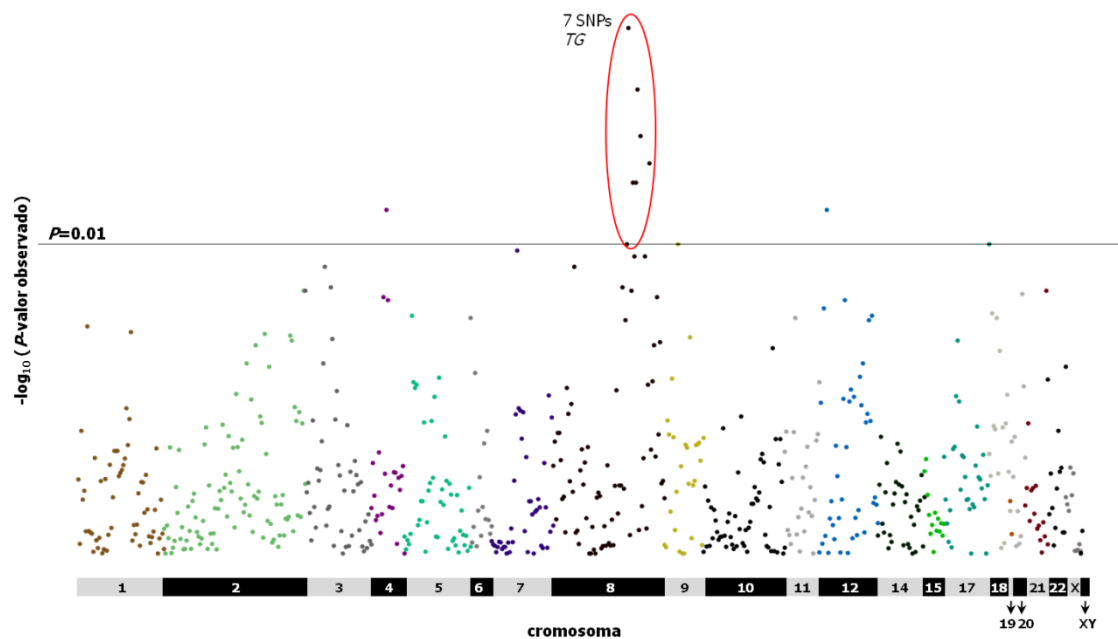


Figura R10. Manhattan plot de la distribución de frecuencias alélicas entre CPTvf y controles

Tabla R5. Resultados muy significativos ($P < 0.01$) al comparar el subtipo CPT clásico frente a controles en la serie 1.

Cr	Coord	SNP	Gen	Posición	Función	Alelos	P^a
9	99664423	rs10119760	FOXE1	downstream	tagSNP	C/G	1×10^{-5}
14	80623486	rs2284734	TSHR	intrónico	tagSNP	A/G	4×10^{-5}
14	80639431	rs1017141	TSHR	intrónico	tagSNP	A/G	1×10^{-4}
9	99653961	rs3758249	FOXE1	upstream	TFBS	A/G	2×10^{-4}
9	99645254	rs7048394	FOXE1	upstream	tagSNP	C/T	2×10^{-4}
9	99652091	rs894673	FOXE1	upstream	tagSNP	A/T	2×10^{-4}
9	99654938	rs907577	FOXE1	upstream	TFBS	A/G	3×10^{-4}
14	80623539	rs2284735	TSHR	intrónico	tagSNP	A/G	1×10^{-3}
9	99656842	rs3021526	FOXE1	exónico	p.Ser275Ser	C/T	2×10^{-3}
14	80644758	rs3783941	TSHR	exónico	p.Arg248Ser	G/T	2×10^{-3}
14	80641412	rs7151769	TSHR	intrónico	tagSNP	C/T	3×10^{-3}

La tabla está ordenada por P -valor (de más a menos significativo); ^a P -valor asociado a la diferencia de frecuencias alélicas; Abreviaturas: Cr= cromosoma; Coord= coordenada; TFBS= *Transcription Factor Binding Sites*.

Tabla R6. Resultados muy significativos ($P < 0.01$) al comparar la variante folicular de CPT (CPTvf) frente a controles en la serie 1.

Cr	Coord	SNP	Gen	Posición	Función	Alelos	P^a
8	133991009	rs2687834	TG	intrónico	tagSNP	A/C	4×10^{-4}
8	134044465	rs2069561	TG	exónico	p.Asp1838Asn	A/G	1×10^{-3}
8	134052230	rs10106773	TG	intrónico	tagSNP	A/G	2×10^{-3}
8	134191878	rs2979033	TG	intrónico	tagSNP	A/G	3×10^{-3}
8	134000930	rs853304	TG	exónico	ESE	A/G	4×10^{-3}
8	134041895	rs2142384	TG	intrónico	tagSNP	A/G	4×10^{-3}
4	55249088	rs2703488	KIT	intrónico	tagSNP	A/G	6×10^{-3}
12	28000415	rs3910902	PTHLH	downstream	tagSNP	A/T	6×10^{-3}
8	133989700	rs2069556	TG	exónico	p.Asp1312Gly	A/G	9×10^{-3}
9	5011514	rs7034753	JAK2	intrónico	TFBS	A/G	9×10^{-3}

La tabla está ordenada por P -valor (de más a menos significativo); ^a P -valor asociado a la diferencia de frecuencias alélicas; Abreviaturas: Cr= cromosoma; Coord= coordenada; ESE= *Exonic Splicing Enhancers*; TFBS= *Transcription Factor Binding Sites*.

2.2.4. Resultados en Carcinoma Folicular de Tiroides (CFT)

Los resultados de la distribución de frecuencias entre casos con CFT y controles se muestran en la tabla R7. Este tipo tumoral, aunque bien definido histológicamente, es bastante infrecuente, y en nuestra serie 1 sólo disponíamos de 69 casos, por lo que deberíamos tratar estos resultados con cautela. Esta comparación señaló varios genes, como el del *peroxisome proliferator-activated receptor gamma* (*PPARG*), el gen de la pendrina (*SLC26A4*), la *Janus* kinasa 2 (*JAK2*) o el gen asociado al adenoma tiroideo (*THADA*). Entre ellos, destacó el gen del receptor de la hormona del crecimiento (*GHR*), con tres variantes asociadas a ese nivel de significación ($P < 0.01$). El gen *GHR* no estaba asociado a CPT, y podría ser específico de CFT (LR $P = 0.002$ para el rs6179 al considerar CPT y CFT).

2.2.5. Resultados en carcinomas con patrón de crecimiento folicular (CPTvf y CFT)

Dado que el subtipo CPTvf presenta características intermedias entre un CPT y un CFT, nos propusimos hacer un análisis considerando juntos los pacientes que hubiesen desarrollado bien CPTvf o CFT. Los resultados, resumidos en la tabla R8, reprodujeron algunas de las asociaciones encontradas para cada subtipo por separado, como los obtenidos para el gen de la tiroglobulina (*TG*) en CPTvf, además de otras variantes en genes asociados a CFT, como *JAK2*, *SLC26A4*, y especialmente *PPARG*. Además, obtuvimos resultados muy significativos para una variante en el gen que codifica para el homólogo de *ras*, miembro G (*RHOG*).

Tabla R7. Resultados muy significativos ($P < 0.01$) al comparar casos de carcinoma folicular de tiroides (CFT) frente a controles en la serie 1.

Cr	Coord	SNP	Gen	Posición	Función	Alelos	P^a
14	80481870	rs12895801	<i>TSHR</i>	<i>upstream</i>	tagSNP	C/G	1×10^{-3}
3	12443589	rs13099828	<i>PPARG</i>	intrónico	tagSNP	C/G	4×10^{-3}
5	42735801	rs6179	<i>GHR</i>	exónico	ESE	A/G	4×10^{-3}
10	51240158	rs10740051	<i>NCOA4</i>	intrónico	tagSNP	A/G	8×10^{-3}
2	43633169	rs1873555	<i>THADA</i>	intrónico	tagSNP	C/T	8×10^{-3}
7	107086763	rs2701684	<i>SLC26A4</i>	<i>upstream</i>	tagSNP	A/G	8×10^{-3}
9	4969730	rs1887427	<i>JAK2</i>	<i>upstream</i>	tagSNP	A/G	9×10^{-3}
2	43631158	rs11903287	<i>THADA</i>	intrónico	tagSNP	A/T	1×10^{-2}
5	42715853	rs4547964	<i>GHR</i>	intrónico	tagSNP	A/G	1×10^{-2}
5	42755583	rs2910875	<i>GHR</i>	3'UTR	ESE	C/T	1×10^{-2}

La tabla está ordenada por P -valor (de más a menos significativo); ^a P -valor asociado a la diferencia de frecuencias alélicas; Abreviaturas: Cr= cromosoma; Coord= coordenada; ESE= *Exonic Splicing Enhancers*; UTR= región no traducida (*untranslated region*).

Tabla R8. Resultados muy significativos ($P < 0.01$) al comparar los carcinomas con patrón de crecimiento folicular (CPTvf+CFT) frente a controles en la serie 1.

Cr	Coord	SNP	Gen	Posición	Función	Alelos	P^a
8	133991009	rs2687834	<i>TG</i>	intrónico	tagSNP	A/C	2×10^{-3}
3	12443589	rs13099828	<i>PPARG</i>	intrónico	tagSNP	C/G	3×10^{-3}
8	134044465	rs2069561	<i>TG</i>	exónico	p.Asp1838Asn	A/G	3×10^{-3}
8	134000930	rs853304	<i>TG</i>	exónico	ESE	A/G	5×10^{-3}
8	134199608	rs10096431	<i>TG</i>	intrónico	tagSNP	A/G	6×10^{-3}
11	3813244	rs1451724	<i>RHOG</i>	intrónico	tagSNP	A/G	7×10^{-3}
7	107086763	rs2701684	<i>SLC26A4</i>	<i>upstream</i>	tagSNP	A/G	1×10^{-2}
8	134191878	rs2979033	<i>TG</i>	intrónico	tagSNP	A/G	1×10^{-2}
9	5011514	rs7034753	<i>JAK2</i>	intrónico	TFBS	A/G	1×10^{-2}

La tabla está ordenada por P -valor (de más a menos significativo); ^a P -valor asociado a la diferencia de frecuencias alélicas; Abreviaturas: Cr= cromosoma; Coord= coordenada; ESE= *Exonic Splicing Enhancers*; TFBS= *Transcription Factor Binding Sites*.

2.2.6. Resultados en carcinomas agresivos

Por último, realizamos una comparación teniendo en cuenta, en lugar del tipo tumoral, la presencia de metástasis a nódulos linfáticos o a otros órganos, tanto en el momento del diagnóstico como las desarrolladas durante el seguimiento del paciente. Los resultados más significativos ($P < 0.01$) de la comparación de estos carcinomas "agresivos" frente a controles están resumidos en la tabla R9.

Como puede comprobarse en la tabla, varios genes ya identificados en otras comparaciones, como *FOXE1*, *TSHR* y *KIT*, aparecían también asociados a este grupo. Esto parece ser más una consecuencia de la predominancia de CPT en este grupo que una verdadera asociación de estos genes con carcinomas más agresivos.

Por el contrario, se encontró una asociación para la variante no sinónima rs25882 en el gen *CSF2*, que codifica una citoquina. Dicha asociación sólo se produjo al considerar estos casos con algún tipo de metástasis, y no cuando se tuvieron en cuenta aquéllos que no se hubiesen diseminado fuera del tiroides (LR $P = 0.0086$).

Tabla R9. Resultados muy significativos ($P < 0.01$) al comparar carcinomas agresivos frente a controles en la serie 1.

Cr	Coord	SNP	Gen	Posición	Función	Alelos	P
9	99664423	rs10119760	<i>FOXE1</i>	downstream	tagSNP	C/G	1×10^{-5}
9	99652091	rs894673	<i>FOXE1</i>	upstream	tagSNP	A/T	2×10^{-5}
9	99653961	rs3758249	<i>FOXE1</i>	upstream	TFBS	A/G	2×10^{-5}
9	99654938	rs907577	<i>FOXE1</i>	upstream	TFBS	A/G	2×10^{-5}
9	99645254	rs7048394	<i>FOXE1</i>	upstream	tagSNP	C/T	1×10^{-4}
9	99656842	rs3021526	<i>FOXE1</i>	exónico	p.Ser275Ser	C/T	1×10^{-3}
4	55249088	rs2703488	<i>KIT</i>	intrónico	tagSNP	A/G	1×10^{-3}
4	55266198	rs3111800	<i>KIT</i>	intrónico	tagSNP	C/T	2×10^{-3}
14	80641412	rs7151769	<i>TSHR</i>	intrónico	tagSNP	C/T	8×10^{-3}
14	80623486	rs2284734	<i>TSHR</i>	intrónico	tagSNP	A/G	9×10^{-3}
5	131439359	rs25882	<i>CSF2</i>	exónico	p.Ile117Thr	C/T	1×10^{-2}

La tabla está ordenada por P -valor (de más a menos significativo); ^a P -valor asociado a la diferencia de frecuencias alélicas; Abreviaturas: Cr= cromosoma; Coord= coordenada; TFBS= *Transcription Factor Binding Sites*.

2.3. Estudio de haplotipos

El análisis de haplotipos se realizó en todos aquellos genes que presentasen algún bloque de desequilibrio de ligamiento (LD) según el Proyecto HapMap, y al menos dos tagSNPs con los que estimar los haplotipos. En total se hizo el análisis, empleando el programa PHASE 2.0, para 57 genes, y se obtuvo alguna asociación estadísticamente significativa ($P < 0.05$) en 23 de ellos.

Las comparaciones de la distribución de haplotipos entre casos y controles nos permitieron confirmar los principales resultados del estudio individual de polimorfismos. Entre ellos, destacan los resultados muy significativos ($P < 0.01$) para la distribución de haplotipos en genes tales como *SMAD7* en CPT, *TSHR* y *FOXE1* en CPT clásico, *TG* en CPTvf y *GHR* en CFT. Además, este análisis permitió detectar diferencias significativas ($P < 0.05$) en 4 genes, no asociados en el estudio individual: *KCNIP3* (que codifica la proteína DREAM, un represor transcripcional de elementos de respuesta al calcio), *TRIP11* (una proteína que interactúa con el receptor de hormonas tiroideas), *PTH1R* (un receptor de la hormona paratiroidea) y *CSGALNACT1* (una acetil-galactosaminiltransferasa). En la Tabla R10 se muestran los resultados del estudio de haplotipos.

Tabla R10. Resultados significativos ($P < 0.05$) de la distribución de haplotipos entre casos y controles de la serie 1.

Cr ^a	Gen_bloque LD ^b	SNPs ^c	P ^d	Grupo de casos ^e	N casos ^f
18	SMAD7_b3	4	0.0018	CPT	520
18	SMAD7_b4	2	0.0106	CPT	520
3	RAF1_b2	7	0.0111	CPT	520
18	SMAD7_b2	2	0.0260	CPT	520
8	CSGALNACT1_b9	5	0.0266	CPT	520
12	STAT2_b1	2	0.0270	CPT	520
2	KCNIP3_b1	7	0.0348	CPT	520
4	KIT_b2	5	0.0404	CPT	520
3	PTH1R_b2	6	0.0435	CPT	520
14	TSHR_b11	8	0.0002	CPT clásico	305
9	FOXE1_b1	7	0.0038	CPT clásico	305
14	TSHR_b10	6	0.0079	CPT clásico	305
2	THADA_b5	6	0.0103	CPT clásico	305
8	TG_b3	2	0.0011	CPTvf	146
8	TG_b15	6	0.0064	CPTvf	146
8	TG_b2	7	0.0111	CPTvf	146
8	TG_b17	5	0.0155	CPTvf	146
2	TRIP12_b1	11	0.0208	CPTvf	146
12	SLC5A8_b1	7	0.0348	CPTvf	146
2	PAX8_b5	6	0.0412	CPTvf	146
8	TG_b4	4	0.0433	CPTvf	146
5	GHR_b5	9	0.0023	CFT	69
2	THADA_b17	8	0.0135	CFT	69
8	FGFR1_b2	4	0.0457	CFT	69
9	JAK2_b2	3	0.0464	CFT	69
11	RHOG_b4	4	0.0021	CPTvf+CFT	215
3	PPARG_b3	3	0.0151	CPTvf+CFT	215
14	TRIP11_b2	3	0.0315	CPTvf+CFT	215
7	SLC26A4_b1	4	0.0447	CPTvf+CFT	215
5	CSF2_b2	2	0.0279	Agresivos	211
18	SMAD4_b1	3	0.0456	Agresivos	211

La tabla está ordenada por P -valor y por grupo de casos considerado. ^a Cromosoma donde está localizado el gen; ^b Gen y bloque de desequilibrio de ligamiento estudiado; los SNPs que definen cada bloque están listados en la tabla S1; ^c Número de polimorfismos que representan cada haplotipo; ^d Significación estadística (tras 10000 permutaciones) de la distribución de haplotipos estimada mediante PHASE para casos y controles; ^e Grupo de de casos en el que se obtuvo la asociación significativa; ^f Número de individuos analizados del grupo correspondiente. Abreviaturas: Cr=cromosoma; CPT=Carcinoma Papilar de Tiroides; CFT=Carcinoma Folicular de Tiroides; CPTc=CPT clásico; CPTvf=CPT, variante folicular.

2.4. Selección de genes de baja penetrancia candidatos para la validación

Los resultados obtenidos tanto en el estudio individual de SNPs como en la comparación de frecuencias de haplotipos, nos permitieron identificar varios genes y variantes candidatos a actuar como *loci* de baja penetrancia en los distintos tipos de carcinomas de tiroides de estirpe folicular. Las asociaciones más consistentes, elegidas para ser replicadas en la fase de validación del estudio (series 2+3), están resumidas en la tabla R11. Para cada variante, se muestran las frecuencias obtenidas en casos y controles, los modelos de herencia que mejor se ajustan y los riesgos asociados, junto con sus *P* valores.

La selección de estos 11 genes se basó en varios criterios, tratando de seleccionar, a la vista de los resultados en esta primera fase, aquellos que parecían más representativos de cada subtipo tumoral. Dentro de los polimorfismos en cada gen significativamente asociados con susceptibilidad a cáncer de tiroides, la selección de las variantes a validar se hizo, bien por ser las más consistentemente asociadas en ese *locus* (el mejor tagSNP), o bien por ser potencialmente funcionales (cambios de aminoácidos, unión a factores de transcripción, etc.).

De la comparación con mayor tamaño muestral, es decir, casos de CPT frente a controles, elegimos 5 variantes en 5 genes. Los resultados relativos a *FOXE1*, el gen más significativamente asociado en nuestro estudio, se tratan aparte, en la sección 3 de Resultados. Además, elegimos sendos tagSNPs intrónicos en los genes *SMAD7* y *KIT*, que mostraron varios polimorfismos significativamente asociados al desarrollo de CPT. También se seleccionó un tagSNP en el *locus* de *PRKAR1A*, gen capaz de reordenarse con *RET* en tumores tiroideos. Por último, elegimos la variante no sinónima p.Met594Ile del factor de transcripción *STAT2*.

Respecto a los casos de CPT clásico, elegimos el mejor tagSNP del gen *TSHR*, para el que obtuvimos resultados altamente significativos en la primera fase. Además, esta variante se localiza en una región potencialmente importante para que se lleve a cabo correctamente el proceso de *splicing*, y es parte de un haplotipo muy significativo para esta comparación. Por otra parte, seleccionamos el cambio de aminoácido p.Pro470Thr del gen *CCDC6* (frecuentemente reordenado con *RET* en tumores tiroideos) para el que, sin llegar a la significación estadística, obtuvimos una tendencia ($P \rightarrow 0.05$) para varios polimorfismos (datos no mostrados).

Cuando consideramos la variante folicular de CPT (CPTvf), se eligió el mejor tagSNP del gen de la tiroglobulina (*TG*), siendo este gen significativo para varios SNPs y en el estudio de haplotipos. Respecto al CFT, seleccionamos un SNP que afectaría potencialmente al *splicing* de *GHR*, mientras que al considerar juntos los carcinomas de patrón folicular (CPTvf+CFT), obtuvimos resultados muy consistentes para el gen *PPARG*, que representamos aquí con su mejor tagSNP.

Por último, de la comparación entre los carcinomas que hubiesen presentado algún tipo de metástasis frente a controles, seleccionamos un cambio de aminoácido (p.Ile117Thr) del gen *CSF2*.

Tabla R11. Genes de baja penetrancia y polimorfismos asociados consistentemente al desarrollo de carcinomas de tiroides de estirpe folicular en la serie 1.

SNP	Gen	Posición	Función	Alelos ^a	MAF en controles; casos	Grupo de casos	Modelo	OR (IC 95%) ^b	P ^c
rs907577	<i>FOXE1</i>	<i>upstream</i>	TFBS	A/G	0.399; 0.482	CPT	codominante	1.38 (1.15-1.64)	3.9 x 10 ⁻⁴
rs16973034	<i>PRKAR1A</i>	<i>downstream</i>	tagSNP	A/G	0.153; 0.108	CPT	dominante	0.63 (0.47-0.84)	1.8 x 10 ⁻³
rs2703488	<i>KIT</i>	intrónico	tagSNP	T/C	0.488; 0.551	CPT	dominante	1.43 (1.07-1.91)	1.5 x 10 ⁻²
rs4939827	<i>SMAD7</i>	intrónico	tagSNP	T/C	0.461; 0.416	CPT	recesivo	0.68 (0.50-0.94)	1.9 x 10 ⁻²
rs2066807	<i>STAT2</i>	exónico	p.Met594Ile	G/C	0.033; 0.054	CPT	dominante	1.69 (1.06-2.69)	2.6 x 10 ⁻²
rs2284734	<i>TSHR</i>	intrónico	tagSNP	A/G	0.290; 0.388	CPT clásico	recesivo	2.64 (1.69-4.13)	1.8 x 10 ⁻⁵
rs1053266	<i>CCDC6</i>	exónico	p.Pro470Thr	G/T	0.494; 0.543	CPT clásico	recesivo	1.37 (0.96-1.97)	8.7 x 10 ⁻²
rs2687834	<i>TG</i>	intrónico	tagSNP	G/T	0.447; 0.565	CPTvf	recesivo	2.28 (1.50-3.46)	1.1 x 10 ⁻⁴
rs6179	<i>GHR</i>	exónico	ESE	G/A	0.356; 0.232	CFT	dominante	0.46 (0.27-0.80)	5.4 x 10 ⁻³
rs13099828	<i>PPARG</i>	intrónico	tagSNP	C/G	0.186; 0.256	CPTvf+CFT	dominante	1.71 (1.22-2.38)	1.6 x 10 ⁻³
rs25882	<i>CSF2</i>	exónico	p.Ile117Thr	T/C	0.185; 0.129	agresivos	dominante	0.69 (0.48-0.99)	4.7 x 10 ⁻²

La tabla está ordenada según el grupo de casos considerados y, dentro de cada grupo, por *P* valor. ^a Alelo mayor/alelo menor (en controles); ^b OR obtenidos al comparar casos frente a controles considerando el alelo polimórfico como el alelo de riesgo. ^c Los *P* valores son los asociados al OR, y están ajustados por género y edad como covariables. Abreviaturas: MAF= *minor allele frequency*; OR= *odds ratio*; IC 95%= intervalo de confianza al 95 %; TFBS= *Transcription Factor Binding Site*; ESE= *Exonic Splicing Enhancers*; CPT=carcinoma papilar de tiroides; CPTvf= variante folicular del CPT; CFT= carcinoma folicular de tiroides.

3. Resultados referidos al gen *FOXE1*

3.1. Varios polimorfismos en el *locus* de *FOXE1* están significativamente asociados al desarrollo de carcinoma papilar de tiroides en población española

De los 690 SNPs analizados en la serie 1 del estudio, las variantes situadas en el *locus* de *FOXE1* fueron las que alcanzaron los resultados más significativos. Los resultados para el mejor tagSNP de este gen fueron un OR (por alelo)=1.47; IC 95%=1.23-1.75; $P=2.4 \times 10^{-5}$. Estos resultados sugerían de forma consistente un papel de *FOXE1* como gen de baja penetrancia (GBP), actuando específicamente en el desarrollo de carcinoma papilar de tiroides (CPT). Estos resultados de asociación están resumidos en la tabla R12. Aunque inicialmente los resultados mostraron una significación mayor en CPTc, no obtuvimos diferencias significativas al comparar, mediante un *likelihood-ratio* test, lo obtenido en CPTc frente a otros subtipos de CPT ($P=0.5502$).

3.2. El análisis de haplotipos identifica una combinación de riesgo significativamente asociada al desarrollo de CPTc

De los seis polimorfismos significativos en *FOXE1*, cinco estaban localizados en el mismo bloque de desequilibrio de ligamiento (bloque LD, figura R11), por lo que fueron empleados para el análisis de haplotipos, que incluyó la región del cromosoma 9 comprendida entre las coordenadas 99648503 y 99668059 (figura R11, tabla R13). Esta aproximación, que se empleó para reforzar los resultados obtenidos en el estudio individual de SNPs, mostró una diferencia significativa en la distribución de haplotipos entre casos de CPT y controles (P asociada a PHASE=0.0147). Nuevamente, las diferencias fueron más acusadas al considerar el subtipo CPTc (P asociada a PHASE=0.0005). De los diez haplotipos de *FOXE1* inferidos en nuestra población, tres representaban al 95 % de los individuos. Los siete haplotipos restantes tenían frecuencias inferiores al 1 % y no se tuvieron en cuenta para análisis posteriores. El estudio detallado de los tres haplotipos mencionados anteriormente (como se muestra en la tabla R13) nos permitió identificar un haplotipo de riesgo significativamente sobrerrepresentado en los pacientes con CPTc (OR=1.66; $P=0.0005$).

Tabla R12. Polimorfismos significativos en el *locus* de *FOXE1* al comparar casos de carcinoma papilar de tiroides frente a controles de la serie 1.

SNP	Coordenada	Localización	Función / predicción	Alelos	Controles / casos genotipados	MAF en controles	MAF en casos CPT	OR ^a (IC 95%)	P
rs7048394	99645254	<i>upstream</i>	tagSNP	C/T	504/520	0.243	0.315	1.46 (1.19-1.78)	2.4 x 10 ⁻⁴
rs894673	99652091	<i>upstream</i>	tagSNP	T/A	504/520	0.400	0.482	1.39 (1.17-1.65)	2.2 x 10 ⁻⁴
rs3758249	99653961	<i>upstream</i>	TFBS	G/A	503/517	0.399	0.483	1.39 (1.17-1.66)	1.9 x 10 ⁻⁴
rs907577	99654938	<i>upstream</i>	TFBS	A/G	503/518	0.399	0.482	1.39 (1.17-1.65)	2.1 x 10 ⁻⁴
rs3021526	99656842	exón 1	p.Ser275Ser	T/C	495/475	0.398	0.469	1.32 (1.11-1.58)	2.1 x 10 ⁻³
rs10119760	99664423	<i>downstream</i>	tagSNP	C/G	504/518	0.344	0.437	1.47 (1.23-1.75)	2.4 x 10 ⁻⁵

^a OR por alelo obtenidos al comparar casos frente a controles considerando el alelo polimórfico como el alelo de riesgo. Abreviaturas: MAF=*minor allele frequency*; OR= *odds ratio*; IC 95%=intervalo de confianza al 95 %; TFBS= *Transcription Factor Binding Site*; CPT=carcinoma papilar de tiroides

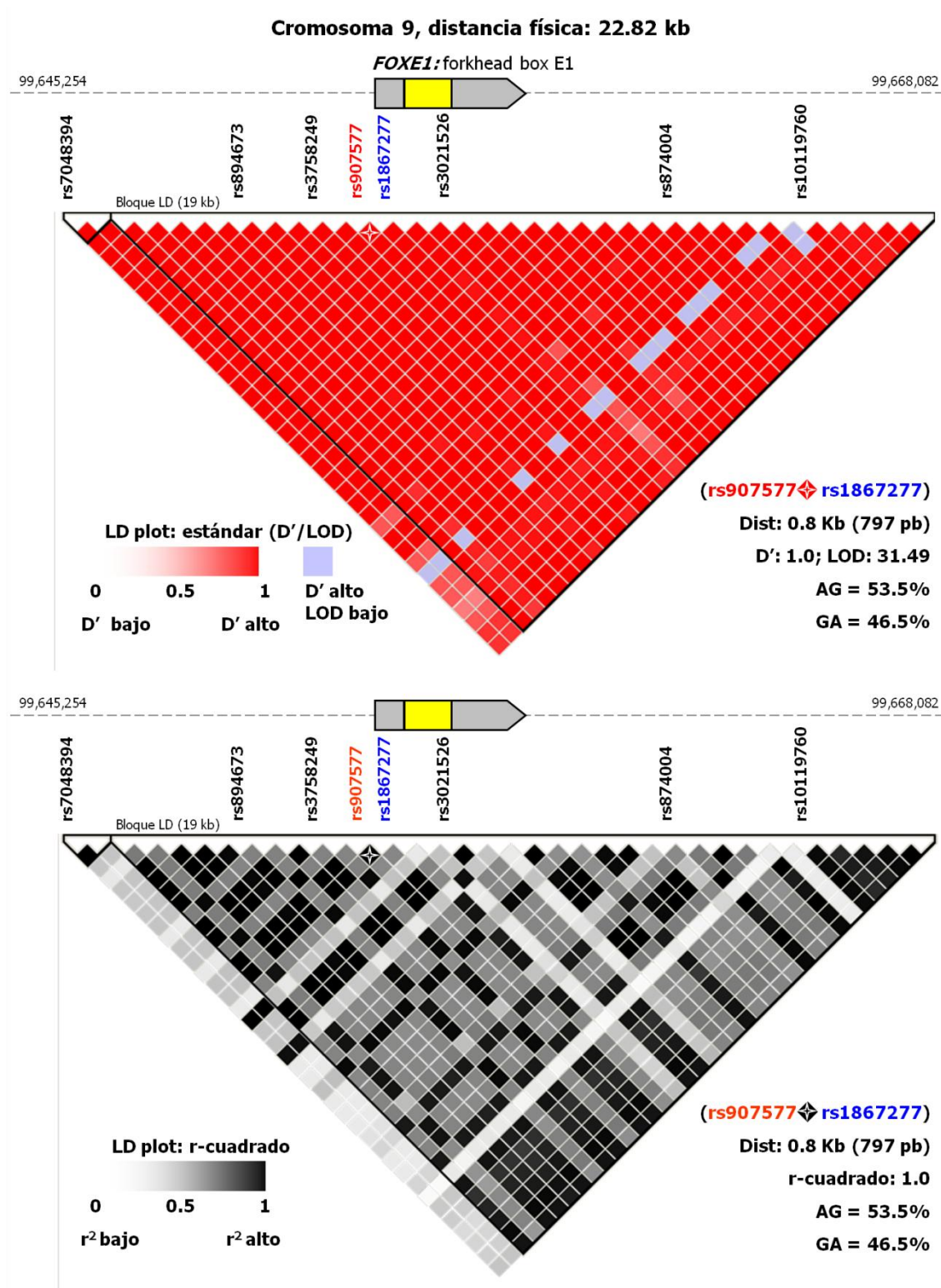


Figura R11. Representación del patrón de desequilibrio de ligamiento (LD plot) en el *locus* de *FOXE1*, basado en los datos de la población HapMap-CEU (*Release 24/phase II Nov08, NCBI B36, dbSNP 126*). La región abarca 22.82 kilobases comprendidas entre las coordenadas 99648503 y 99668059. El gen *FOXE1* está representado en amarillo (exón) y gris (5' y 3' UTR). El LD está expresado con los parámetros D'/LOD (panel superior) y r -cuadrado (panel inferior). Valores de D' y r -cuadrado iguales a 1 significan LD completo. Esta región está capturada por los SNPs que aparecen en la figura, todos ellos altamente significativos en la serie 1 del estudio. La variante funcional rs1867277 aparece en azul, y el tagSNP rs907577, en rojo. Se muestran los valores específicos de LD entre estos dos SNPs (estrella de cuatro puntas).

3.3. Un análisis detallado del haplotipo de riesgo en CPT identifica al SNP rs1867277 como una variante altamente correlacionada y potencialmente funcional.

Dado que ninguna de las seis variantes asociadas en *FOXE1* mostró una predicción funcional consistente (según las herramientas bioinformáticas empleadas), se realizó un análisis más detallado de la secuencia en este bloque LD, con la intención de buscar otras variantes que pudiesen explicar desde un punto de vista funcional el riesgo a desarrollar CPT. De esta forma, identificamos otro SNP funcionalmente interesante (rs1867277, c.-283G>A), que no había sido incluido inicialmente en el ensayo de *Illumina*. De acuerdo con las herramientas bioinformáticas, esta variante podría influir en la unión de factores de transcripción que regularían la transcripción de *FOXE1*. Los resultados obtenidos mediante DHPLC en un subgrupo de 200 individuos nos permitieron confirmar experimentalmente el desequilibrio de ligamiento completo entre dicha variante potencialmente funcional (rs1867277) y el tagSNP rs907577 (incluido en el ensayo *Illumina*), y de esta forma, poder imputar los genotipos de la primera variante (OR por alelo=1.39; $P=2.1 \times 10^{-4}$). Esta correlación está representada en la figura R11, y los valores específicos de LD pueden encontrarse en la tabla S2.

Es interesante señalar que el haplotipo de riesgo identificado (tabla R13, haplotipo #2), contiene los alelos rs907577G y rs1867277A (funcional y localizado a sólo 797 bases *downstream*), que tiene un efecto en la transcripción de *FOXE1* (ver más adelante).

3.4. La asociación significativa para el rs1867277 de *FOXE1* en carcinoma papilar de tiroides se valida en la segunda fase del estudio

Los resultados de asociación obtenidos en población española para la variante rs1867277, localizada en el promotor de *FOXE1*, fueron validados en la serie 2 del estudio, consistente en un grupo independiente de individuos procedentes de Italia. Se comprobó que este polimorfismo también estaba significativamente sobrerrepresentado en un grupo de 405 casos italianos de CPT frente a 525 controles (OR por alelo=1.64; IC 95%=1.31-2.05; P ajustada= 1.3×10^{-5}). Los resultados obtenidos en las series 1 y 2, y la combinación de ambas, están resumidos en la tabla R14. El análisis conjunto para el rs1867277, que incluyó un número total de 984 CPT frente a 1028 controles, confirmó la participación de esta variante de *FOXE1* en el desarrollo de CPT (OR por alelo=1.49; IC 95%=1.30-1.70; P ajustada= 5.9×10^{-9}).

Tabla R13. Análisis de haplotipos en el *locus* de *FOXE1* en casos de carcinoma papilar de tiroides clásico (CPTc) frente a controles de la serie 1

Haplotipo	Frecuencias haplotípicas ^a		SNP (coordenada cromosómica) ^b							Riesgos asociados ^c	
	Controles	Casos (CPTc)	rs894673 (99,652,091)	rs3758249 (99,653,961)	rs907577 (99,654,938)	rs1867277 (99,655,735)	rs3021526 (99,656,842)	rs874004 (99,661,939)	rs10119760 (99,664,423)	OR	P
Haplotipo #1 >	0.26	0.19	T	G	A	G	T	C	C	1	
Haplotipo #2 >	0.34	0.44	A	A	G	A	C	C	G	1.66	0.0005
Haplotipo #3 >	0.34	0.32	T	G	A	G	T	G	C	1.25	0.1355

^a Las frecuencias haplotípicas se estimaron con el programa PHASE, empleando 10000 permutaciones. ^b La región de LD estudiada abarca 19.5 kb entre las coordenadas 99,648,503 a 99,668,059 del cromosoma 9. ^c Los riesgos asociados a cada haplotipo se calcularon con el programa Haplo.Stats en R, considerando el haplotipo sobrerrepresentado en controles frente a casos (haplotipo #1) como referencia. El haplotipo que confiere un riesgo estadísticamente significativo a desarrollar CPTc (haplotipo #2) está resaltado en negrita. Abreviaturas: OR= *odds ratio*; CPTc=carcinoma papilar de tiroides clásico.

Tabla R14. Resultados de asociación combinados para la variante rs1867277 (c.-283G>A) en carcinomas papilares de tiroides frente a controles

Muestras genotipadas (n CPT/ n controles)	Frecuencia del rs1867277A		Resultados sin ajustar				Resultados ajustados ^g	
	CPT	Controles	OR het ^d (IC 95%)	OR hom ^e (IC 95%)	OR por alelo (IC 95%)	<i>P</i> ^f	OR por alelo (IC 95%)	<i>P</i> ^f
Serie 1 - España (518/503) ^a	0.482	0.399	1.46 (1.10-1.94)	1.90 (1.34-2.71)	1.39 (1.17-1.65)	2.1 x 10 ⁻⁴	1.38 (1.15-1.64)	3.9 x 10 ⁻⁴
Serie 1 ampliada (579 ^b /503)	0.484	0.399	1.50 (1.14-1.97)	1.95 (1.38-2.74)	1.41 (1.19-1.67)	8.5 x 10 ⁻⁵	1.39 (1.17-1.65)	1.8 x 10 ⁻⁴
Serie 2 - Italia (405/525) ^c	0.558	0.456	1.64 (1.18-2.30)	2.38 (1.62-3.50)	1.54 (1.27-1.87)	9.9 x 10 ⁻⁶	1.64 (1.31-2.05)	1.3 x 10 ⁻⁵
Series 1+2 (984/1028)	0.515	0.428	1.48 (1.20-1.83)	2.01 (1.56-2.59)	1.42 (1.25-1.61)	4.2 x 10 ⁻⁸	1.49 (1.30-1.70)	5.9 x 10 ⁻⁹

^a Los genotipos del rs1867277 fueron inferidos a partir del tagSNP adyacente rs907577, tras comprobar el LD total entre ambas variantes (ver apartado 3.3).

^b Se genotiparon 61 CPT adicionales para el rs1867277 mediante la tecnología KASPar (ver figura S1 para más detalles). ^c En la serie 2 (validación), se genotipó el SNP rs1867277 mediante sondas KASPar. ^d OR de heterocigotos: GG vs. GA. ^e OR de homocigotos: GG vs. AA. ^f Los *P* valores son los asociados al OR por alelo. ^g Ajustados por edad, género y origen de las muestras. Abreviaturas: OR=*odds ratio*; IC 95%=intervalo de confianza al 95 %; CPT=carcinoma papilar de tiroides.

3.5. El factor de transcripción DREAM se une al promotor de *FOXE1* en la región donde se localiza la variante de riesgo rs1867277

La variante rs1867277 está localizada en una región muy similar a la secuencia consenso de unión de factores DRE ("downstream regulatory elements"), caracterizada en el promotor del gen de la prodinorfina (*PDYN*) (Jacobson, et al., 2006) (figura R12).

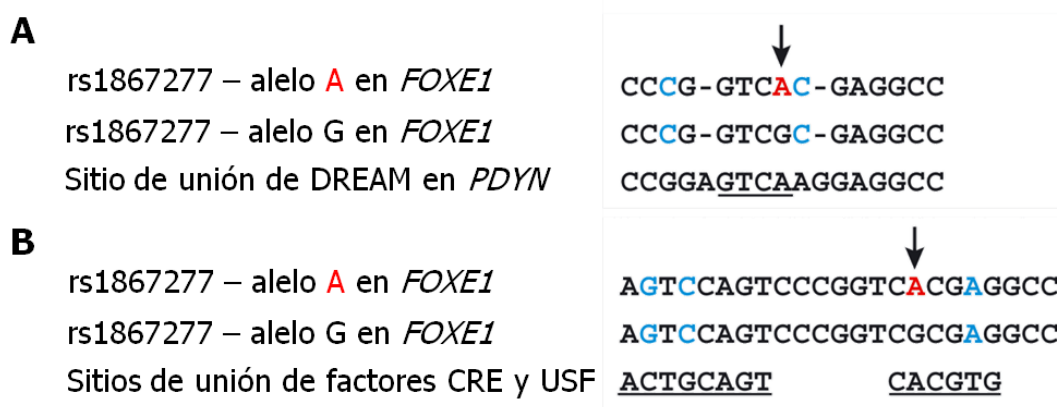


Figura R12. Análisis de la secuencia de la región del promotor de *FOXE1* donde se encuentra la variante rs1867277. **A.** El alelo de riesgo A está resaltado en rojo e indicado con una flecha. Los nucleótidos en azul representan los desemparejamientos respecto a la secuencia consenso de unión de DREAM. Los espacios (indicados con guiones), se introdujeron para optimizar el alineamiento con la secuencia de DREAM. El sitio de unión de DRE en el promotor del gen de la prodinorfina humana (*PDYN*) está subrayado. **B.** Alineamiento de las secuencias próximas al polimorfismo. Las secuencias consenso de unión de los factores de transcripción CRE y USF están subrayadas, mostrándose en azul los desemparejamientos de la secuencia del promotor de *FOXE1* con ellas.

En un primer intento para evaluar el papel de la variante rs1867277 en la regulación transcripcional de *FOXE1*, se evaluó, en extractos nucleares de la línea celular WRO, la capacidad de unión de los alelos G y A del SNP, mediante un ensayo de retardo (EMSA). Como se muestra en la figura R13, se formaron diferencialmente complejos de ADN-proteína al usar sondas de los alelos A y G (calles 2 y 5, respectivamente). Se formó un complejo inferior en presencia de ambos alelos, mientras que un complejo superior se formó exclusivamente en presencia del alelo de riesgo rs1867277A. Se comprobó que la formación de los complejos era específica, al observar una disminución de la intensidad de la banda en el gel al realizar ensayos de competición con un exceso (100X) de sondas de secuencia específica pero no marcadas (calles 3 y 6), pero no con una secuencia no relacionada (calles 4 y 7).

A continuación, intentamos identificar los factores que se unen a ambos alelos, centrándonos primero en DREAM ("Downstream Regulatory Element Antagonist Modulator"), ya que el rs1867277 está localizado en un posible sitio consenso de unión de elementos DRE (figura R12). Utilizando una sonda con la secuencia del promotor de la prodinorfina, que contiene la secuencia consenso de unión a DRE (DRE-pDYN), observamos la formación de un complejo con la misma movilidad electroforética que el complejo inferior formado con los alelos A y G (figura R13, calle 8; competición con sonda específica y no relacionada en calles 9 y 10). Curiosamente, observamos una competición parcial por este complejo al emplear oligos con las secuencias de los alelos A y G (calles 11 y 12). Además, el empleo de un anticuerpo específico anti-DREAM redujo sustancialmente la intensidad de esta banda inferior (calle 14).

En resumen, estos datos demuestran la participación del factor DREAM endógeno en el complejo inferior, aunque la falta de competición total de los oligos con la secuencia de los alelos A y G sugiere que otras proteínas podrían ser parte de dicho complejo. Es especialmente interesante destacar que se detectó una secuencia de unión a CRE cercana al rs1867277 (figura R12), y se sabe que los factores CRE pueden interactuar conjuntamente con DREAM (Ledo, et al., 2000).

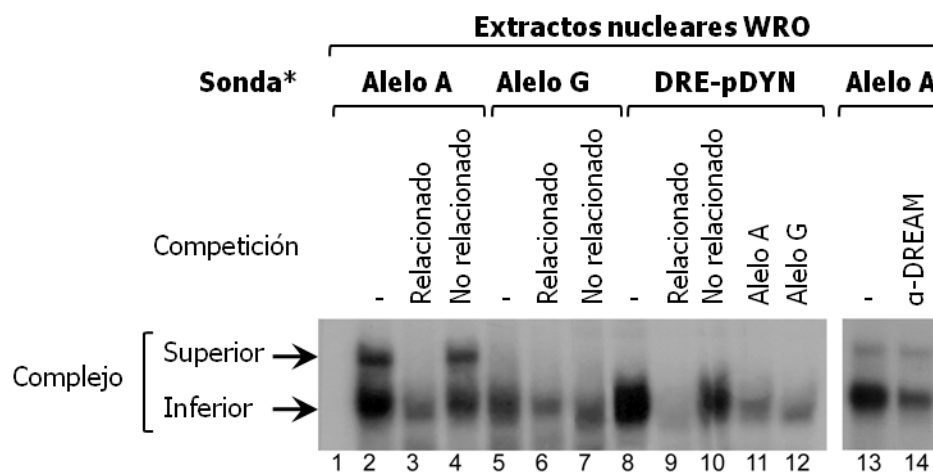


Figura R13. Ensayos de retardo para evaluar la unión de proteínas nucleares a la región del polimorfismo rs1867277 de *FOXE1*. Se incubaron extractos nucleares de células WRO con oligos marcados con ^{32}P ("sonda*"), con las secuencias de rs1867277A, rs1867277G, o el consenso de unión a DRE del promotor de la prodinorfina (DRE-pDYN). Las sondas se incubaron sin (calle 1) o con extractos nucleares (calles 2-14). Los ensayos de competición se realizaron con un exceso molar 100X de sondas no marcadas de secuencias específicas (calles 3, 6, 9), no relacionadas (calles 4, 7, 10) o con los alelos A y G (calles 11 y 12), y también con un anticuerpo anti-DREAM (calle 14). Los complejos ADN-proteína se indican mediante flechas.

3.6. Las proteínas USF se unen específicamente al alelo A del rs1867277

Dado que el complejo superior de ADN-proteína se formaba exclusivamente en presencia del alelo de riesgo rs1867277A, decidimos tratar de identificar el factor o factores de transcripción que formaban parte de ese complejo. Según la base de datos *Gene Regulation* (<http://www.gene-regulation.com>), las proteínas USF1 y USF2 ("*Upstream Stimulatory Factors*") se unirían exclusivamente en presencia del alelo A.

Por tanto, se realizaron ensayos de retardo empleando una secuencia que contenía el alelo A, y se utilizaron como competidores varias sondas u oligos no marcados, incluyendo la secuencia consenso de USF, y varias secuencias de unión de factores CRE, dada la relación entre los elementos CRE y DRE (figura R14A). En este experimento, cuando el oligo con la secuencia del alelo G se utilizó como competidor, sólo disminuyó la intensidad de banda del complejo inferior, demostrándose así la especificidad del complejo superior por el alelo A (figura R14A, calle 5). De forma similar, la secuencia similar a CRE en *FOXE1*, la secuencia consenso de CRE y la secuencia consenso DRE-pDYN, compitieron por la unión en la banda inferior (calles 6-8), sugiriendo que el complejo inferior contenía más de un factor de transcripción. Además, fue interesante comprobar que la competición con la secuencia consenso de USF era capaz de hacer desaparecer completamente la banda superior (calle 9). Además, al emplear las proteínas USF1 y USF2, producidas *in vitro* mediante el sistema TNT, se formó un complejo con la misma movilidad que el complejo superior (calles 11 y 12). La implicación de los factores USF se comprobó mediante el uso de anticuerpos específicos anti-USF1 y anti-USF2, que produjeron un gran retardo o *supershift* del complejo superior, tanto en presencia de extractos proteicos nucleares (figura R14B, calles 6 y 14), como empleando específicamente proteínas USF traducidas *in vitro* (calles 8 y 16). Esto demostraba inequívocamente que USF1/USF2 forman parte del complejo superior. Por otra parte, ensayos de retardo similares realizados con el alelo rs1867277G, no produjeron este complejo superior de ADN-proteína (no mostrado).

En conjunto, estos resultados indican de forma sólida que sólo el alelo rs1867277A es capaz de formar un complejo proteico que incluye a los factores de transcripción USF1 y USF2. Además, ambos alelos (A y G) del rs1867277 de *FOXE1* forman un complejo en el que DREAM, y posiblemente otros factores de transcripción relacionados con CRE, participan.

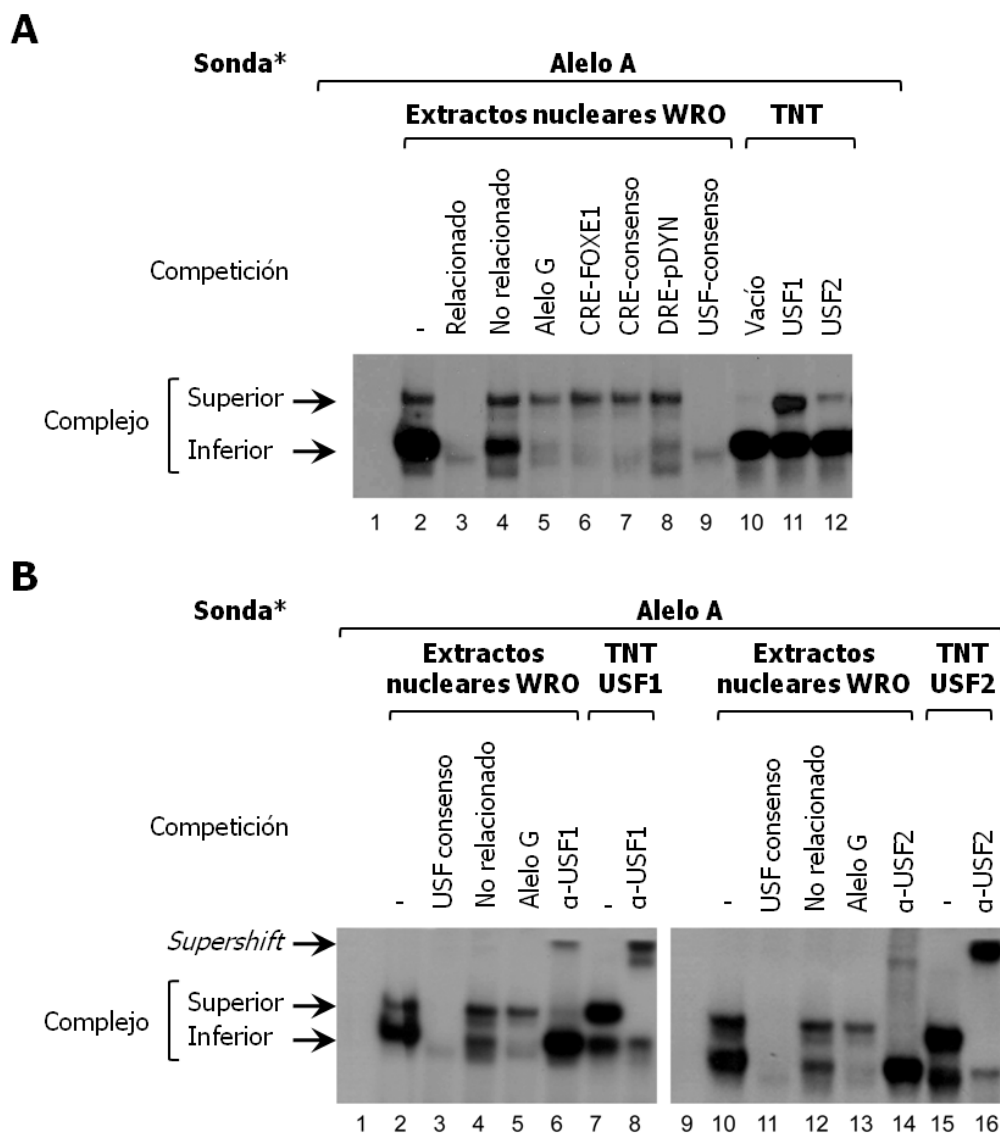


Figura R14. Ensayos de retardo para evaluar la unión de proteínas nucleares a la región del polimorfismo rs1867277 de *FOXE1* (II). **A.** Se empleó una sonda con la secuencia del alelo rs1867277A, marcada con ^{32}P ("sonda*"), y se incubó sola (calle 1), con extractos nucleares de WRO (calles 2-9) o con proteínas producidas *in vitro* mediante el sistema TNT, a partir de un vector vacío (calle 10), o de los vectores de expresión de USF1 y USF2 (calles 11 y 12). Los ensayos de competición se realizaron con un exceso molar 100X de sondas de oligos no marcados de secuencias relacionada (calle 3), no relacionada (calle 4), la secuencia del rs1867277G (calle 5), la secuencia de CRE en el promotor de *FOXE1* ("CRE-FOXE1", calle 6), el consenso de CRE ("CRE-consenso", calle 7), DRE-pDYN (calle 8) y el consenso de USF (calle 9). **B.** La sonda* del rs1867277A se incubó sola (calles 1 y 9), con extractos nucleares de WRO (calles 2-6 y 10-14) o con proteína TNT-USF1 (calles 7 y 8) o TNT-USF2 (calles 15 y 16). La competición se llevó a cabo con un exceso molar 100X de sondas no marcadas de secuencias USF consenso (calles 3 y 11), no relacionadas (calles 4 y 12) o rs1867277G (calles 5 y 13). Para los ensayos de *supershift* se usaron anticuerpos anti-USF1 (calles 6 y 8) o anti-USF2 (calles 14 y 16). Los complejos ADN-proteína se indican mediante flechas.

3.7. Los factores USF1 y USF2 regulan la actividad transcripcional del promotor de *FOXE1*

Para validar el significado funcional de los factores de transcripción identificados sobre la expresión de *FOXE1*, se realizaron ensayos para determinar el papel de dichos factores en la actividad del promotor. Para ello, transfectamos de forma transitoria células HeLa con vectores de expresión que portaban ADNc de los distintos factores de transcripción, y se evaluaron con la construcción reportera que contenía el alelo rs1867277A del promotor de *FOXE1* (pGI3b-FOX E1-283A). Mientras que la transfección con DREAM no produjo variaciones significativas en la actividad del promotor de *FOXE1*, la transfección con USF1 y USF2 incrementó su actividad ($P<0.01$ y $P<0.05$, respectivamente, figura R15A). Dado que en la literatura está descrita la formación de homo- y heterodímeros de USF (Viollet, et al., 1996), cotransfectamos ambos vectores de expresión de forma simultánea, obteniendo un gran incremento (8 veces, $P<0.0001$) en la actividad del promotor de *FOXE1*. Se demostró que este efecto era específico, ya que una forma dominante negativa de USF (DN-USF) era capaz de abolirlo. Además, la cotransfección del vector de expresión de DREAM junto a USF1/USF2, redujo claramente la activación de *FOXE1*, comparada con la condición USF1/USF2 por sí sola ($P<0.0001$, figura R15A), confirmando la función como correpresor descrita para DREAM (Carrion, et al., 1999; Rivas, et al., 2004). Todos estos datos resaltan el importante papel regulador de los factores USF en la expresión de *FOXE1*, así como el efecto inhibitor de DREAM en la activación de *FOXE1* dependiente de USF.

Para verificar que el efecto de los factores USF está mediado por el mencionado sitio de unión en el promotor de *FOXE1* (condicionado por la presencia del alelo rs1867277A), también realizamos transfecciones transitorias empleando la construcción pGI3b-FOX E1-283G (que contiene el alelo rs1867277G). Como se muestra en la figura R15A, no se observó ningún efecto cuando USF1 y USF2 se emplearon de forma independiente, mientras que la cotransfección simultánea de USF1/2 resultó en una activación transcripcional claramente menor que la observada con el alelo A ($P<0.0001$). En resumen, el alelo rs1867277G afecta parcialmente el reclutamiento de los factores USF1/2 y su unión al promotor de *FOXE1*, y altera por tanto la expresión del gen *FOXE1*.

3.8. Las isoformas α CREM y CREB activan la transcripción de *FOXE1* mediante un sitio CRE localizado cerca del rs1867277

La actividad transcripcional del gen *FOXE1* está regulada mediante factores hormonales, particularmente por la tirotropina o TSH, vía AMP cíclico (AMPc) (Ortiz, et al., 1997). Los factores de transcripción α CREM y CREB se unen a secuencias consenso CRE, localizadas en los promotores de genes regulados por AMPc. Dada la existencia de un sitio de unión similar a CRE en la región del promotor de *FOXE1* cercana al SNP rs1867277 (figura R12), estudiamos el papel de α CREM y CREB en la regulación de este gen. La sobreexpresión de ambas isoformas indujeron aumentos significativos de la actividad del promotor de *FOXE1*, siendo α CREM el factor activador más potente (cambios de más de 50 veces, figura R15B, panel izquierdo).

Para investigar si esta activación dependiente de α CREM podría ser modulada por otros factores que se unen a la región del rs1867277 en el promotor de *FOXE1*, se transfectaron, solos o en combinación, los vectores de expresión de USF1, USF2, CREB y DREAM, junto con una cantidad fija de α CREM y la construcción reportera de *FOXE1* (figura R15B, panel derecho). Observamos que ni USF1/USF2 ni CREB afectaban a la actividad del promotor de *FOXE1* inducida por α CREM. Por el contrario, la sobreexpresión de DREAM produjo una represión de la actividad del 80 % ($P < 0.0001$). Teniendo en cuenta estos resultados, podemos concluir que los factores de unión a motivos CRE actúan como reguladores positivos de la expresión de *FOXE1*, mientras que DREAM muestra un efecto regulador negativo sobre la activación producida por α CREM. Por su parte, los factores USF no modulan la actividad transcripcional regulada por α CREM, lo cual sugiere la existencia de dos mecanismos independientes de regulación del promotor de *FOXE1*: uno dependiente de los factores de unión al motivo CRE, y otro determinado por los factores USF, con el represor DREAM regulando ambos mecanismos.

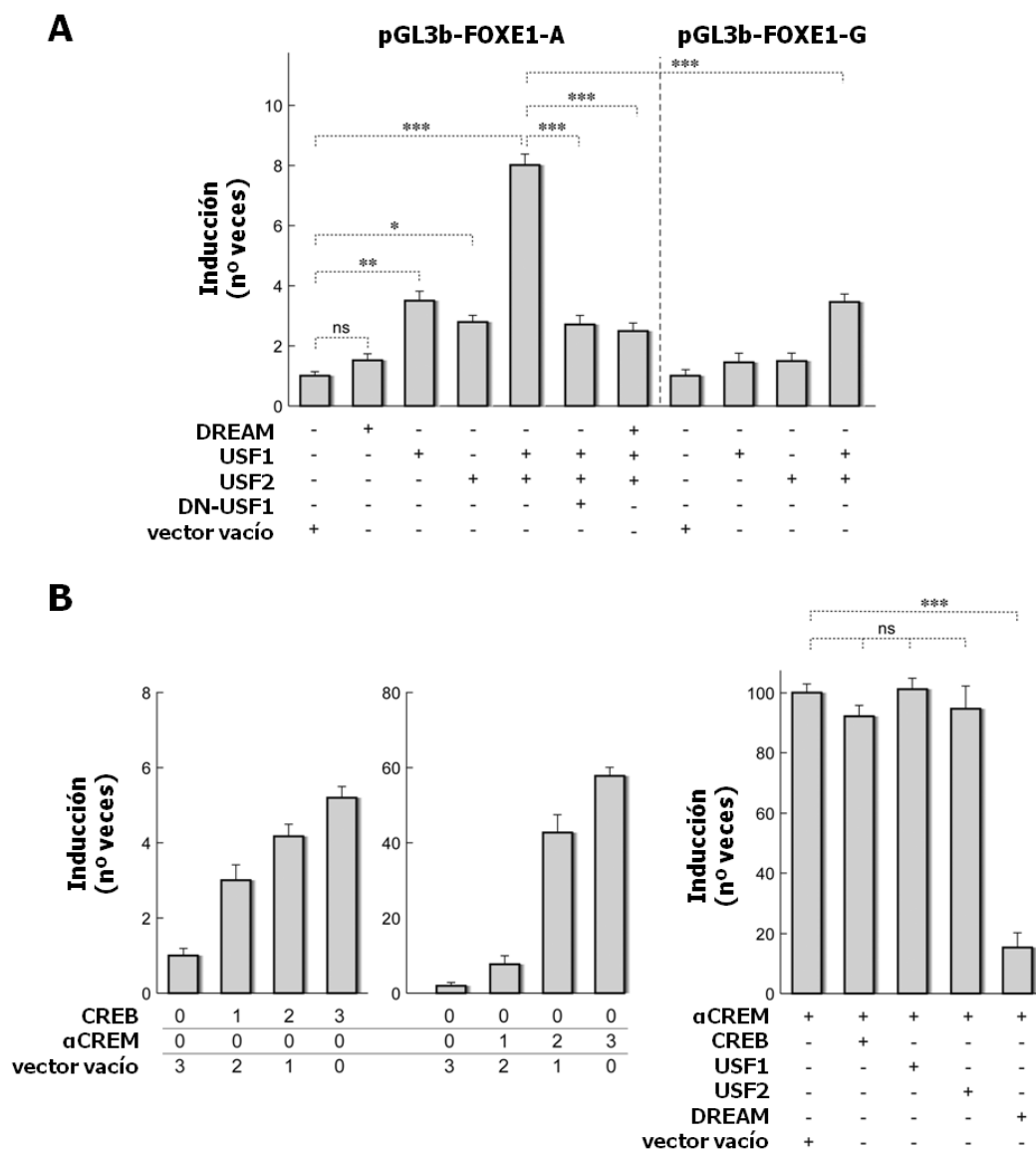


Figura R15. Actividad transcripcional del promotor de *FOXE1*. **A.** Se cotransfectaron células HeLa con la construcción reportera del promotor de *FOXE1*, conteniendo el alelo rs1867277A (pGL3b-FOXE1-283A), el alelo rs1867277G (pGL3b-FOXE1-283G) o el vector vacío, junto a los vectores de expresión con los ADNc, según el caso, de DREAM, USF1, USF2 o una forma dominante negativa de USF1 (DN-USF1). Se emplearon 3 µg de cada vector, salvo en la cotransfección de vectores USF1/USF2, para la que se emplearon 1.5 µg de cada uno. La actividad del promotor está expresada como inducción, en número de veces, respecto a la actividad observada con el vector vacío (unidades relativas). **B.** Se cotransfectaron células HeLa con la construcción reportera de *FOXE1* rs1867277A (pGL3b-FOXE1-283A), y vectores de expresión de CREB, αCREM, USF1, USF2, DREAM o el vector vacío, combinados como se muestran en la figura. Los números en el panel izquierdo representan los µg empleados de cada vector, mientras que en el panel derecho se emplearon 3 µg de los vectores indicados. La actividad del promotor está expresada como inducción, en número de veces (unidades relativas), respecto a la actividad de los vectores vacío (panel izquierdo) o αCREM (panel derecho).

La actividad luciferasa se normalizó con la actividad Renilla derivada de la cotransfección con el vector pRL-Tk, para ajustar por la eficiencia de la transfección. Los resultados son la media±DE de cuatro experimentos independiente. Abreviaturas: (*): $P < 0.05$; (**): $P = 0.01-0.001$; (***) : $P < 0.001$; ns, no significativo.

4. Resultados de validación de variantes seleccionadas del SNP array: segunda fase del estudio de asociación

4.1. Resultados técnicos del genotipado en las series 2 y 3

La tasa de éxito del genotipado de los 11 SNPs en la fase de validación fue ligeramente menor de lo habitual, con unos porcentajes medios de muestras con genotipo del 93.6 % y el 94.5 % para controles y casos, respectivamente. Esto se debió, probablemente, a la baja calidad del ADN genómico disponible. Por tanto, la elección de la plataforma KASPar, que permite trabajar con ADN menos concentrado y de peor calidad, parece que fue muy adecuada para esta fase del estudio. La figura R16 muestra el número de muestras consideradas para cada polimorfismo y la tasa de error en el genotipado.

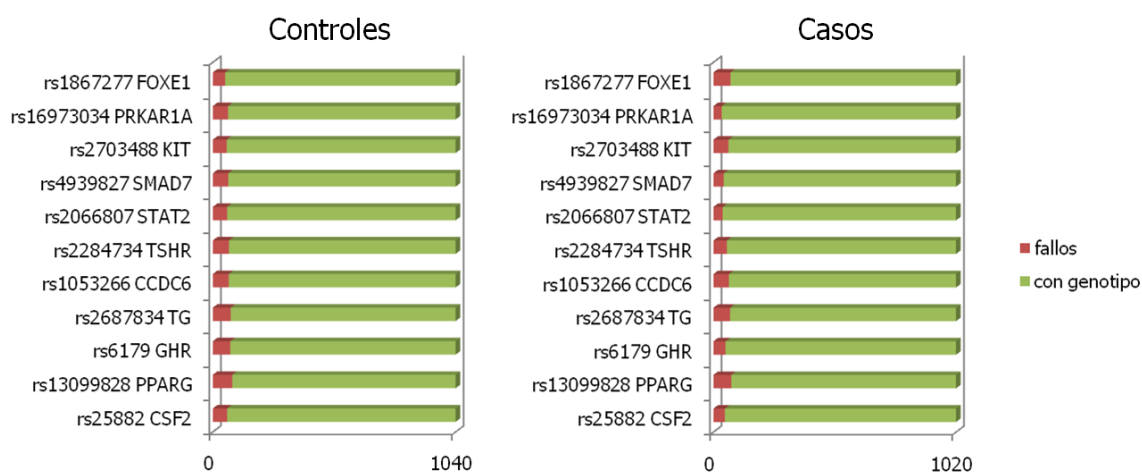


Figura R16. Tasas de error en el genotipado de los 11 SNPs elegidos en la fase de validación

Por otra parte, la especificidad del genotipado fue óptima, obteniendo una concordancia del 100 % en los duplicados intra- e interplaca, y una ausencia total de señal en todos los blancos incluidos en el estudio.

4.2. Resultados de validación de polimorfismos

Una vez comprobada la calidad y especificidad del genotipado mediante la tecnología KASPar, realizamos los análisis apropiados para comprobar si las 11 variantes elegidas en la primera fase del estudio se validaban en esta segunda fase. A continuación, y en las tablas R15 y R16, se muestran los resultados de asociación para cada grupo de casos *versus* controles.

4.2.1. Resultados en Carcinoma Papilar de Tiroides (CPT)

Para la comparación de las cinco variantes asociadas a CPT en serie 1, en la serie de validación disponíamos de 863 casos diagnosticados como CPT, además de 1040 controles. En primer lugar, comprobamos que la variante rs1867277 de *FOXE1*, ampliamente tratada en la sección anterior, también estaba significativamente asociada al desarrollo de CPT al considerar la serie 3 de casos y controles (antes no disponible).

Por el contrario, los tagSNPs en los *loci* de *PRKAR1A* y *KIT* no se validaron ($P > 0.05$), lo cual puede significar que estos genes en realidad no actúan como GBP, o por el contrario, que estos tagSNPs no capturan adecuadamente las respectivas variantes funcionales de cada gen. La variante exónica de *STAT2* tampoco se validó, posiblemente debido a su bajo MAF (aproximadamente un 4 %), lo que limita la potencia del estudio para este SNP concreto.

Por último, para el tagSNP rs4939827 de *SMAD7*, obtuvimos resultados significativos, pero en dirección opuesta a lo esperado. Así, mientras que los portadores del genotipo polimórfico CC de la serie 1 mostraron una protección significativa a desarrollar CPT, los portadores del mismo genotipo en las series 2+3, mostraban riesgo. Un análisis más detallado de las series 2 y 3 por separado demostró que dicha variante no estaba asociada en la serie 2 (OR=1.09; IC 95%=0.75-1.58; $P=0.650$), pero confería riesgo en la serie 3 (OR=1.53; IC 95%=1.10-2.12; $P=0.011$). La implicación que tienen estos resultados se discute en el apartado correspondiente.

Tabla R15. Resultados obtenidos en la fase de validación (series 2+3) para las 11 variantes candidatas.

SNP	Gen	Posición	Función	Alelos	MAF en controles; casos	Grupo de casos	Modelo	OR ^a (IC 95%)	P ^a	OR ^b (IC 95%)	P ^b
rs1867277	FOXE1	5'UTR	TFBS	G/A	0.416; 0.518	CPT	codominante	1.50 (1.31-1.71)	2.2 x 10 ⁻⁹	1.49 (1.29-1.72)	3.8 x 10 ⁻⁸
rs16973034	PRKAR1A	downstream	tagSNP	A/G	0.138; 0.133	CPT	dominante	1.01 (0.82-1.25)	0.917	1.00 (0.79-1.26)	0.984
rs2703488	KIT	intrónico	tagSNP	T/C	0.476; 0.472	CPT	dominante	0.93 (0.76-1.14)	0.468	0.93 (0.75-1.16)	0.522
rs4939827	SMAD7	intrónico	tagSNP	T/C	0.438; 0.478	CPT	recesivo	1.23 (0.98-1.55)	0.073	1.32 (1.03-1.68)	0.027
rs2066807	STAT2	exónico	Met594Ile	G/C	0.040; 0.039	CPT	dominante	0.96 (0.68-1.35)	0.800	0.98 (0.68-1.42)	0.928
rs2284734	TSHR	intrónico	tagSNP	A/G	0.304; 0.317	CPT clásico	recesivo	1.27 (0.90-1.79)	0.176	1.42 (0.99-2.03)	0.058
rs1053266	CCDC6	exónico	Pro470Thr	G/T	0.471; 0.502	CPT clásico	recesivo	1.25 (0.98-1.60)	0.070	1.24 (0.96-1.60)	0.103
rs2687834	TG	intrónico	tagSNP	G/T	0.491; 0.477	CPTvf	recesivo	0.75 (0.48-1.15)	0.187	0.74 (0.46-1.17)	0.195
rs6179	GHR	exónico	ESE	G/A	0.299; 0.286	CFT	dominante	0.96 (0.63-1.46)	0.853	0.87 (0.55-1.37)	0.543
rs13099828	PPARG	intrónico	tagSNP	C/G	0.168; 0.189	CPTvf+CFT	dominante	0.92 (0.66-1.29)	0.640	1.01 (0.71-1.46)	0.936
rs25882	CSF2	exónico	Ile117Thr	T/C	0.194; 0.214	agresivos	dominante	1.01 (0.70-1.46)	0.949	1.08 (0.74-1.57)	0.689

OR obtenidos al comparar casos frente a controles considerando el alelo polimórfico como el alelo de riesgo. ^a Resultados sin ajustar; ^b Resultados ajustados por las covariables edad, género y origen. Abreviaturas: MAF= *minor allele frequency*; OR= *odds ratio*; IC 95%= intervalo de confianza al 95 %; TFBS= *Transcription Factor Binding Site*; ESE= *Exonic Splicing Enhancers*; CPT=carcinoma papilar de tiroides; CPTvf= variante folicular del CPT; CFT= carcinoma folicular de tiroides; UTR= región no traducida (*untranslated region*).

Tabla R16. Comparación de los resultados obtenidos en las fases de descubrimiento (serie 1) y de validación (series 2+3) para las 11 variantes candidatas.

SNP	Gen	Grupo de casos	Modelo	Resultados en la serie 1			Resultados en las series 2+3 ^a		
				MAF en controles; casos	OR (IC 95%)	P	MAF en controles; casos	OR (IC 95%)	P
rs1867277	<i>FOXE1</i>	CPT	codominante	0.399; 0.482	1.38 (1.15-1.64)	3.9×10^{-4}	0.416; 0.518	1.49 (1.29-1.72)	3.8×10^{-8}
rs16973034	<i>PRKAR1A</i>	CPT	dominante	0.153; 0.108	0.63 (0.47-0.84)	1.8×10^{-3}	0.138; 0.133	1.00 (0.79-1.26)	0.984
rs2703488	<i>KIT</i>	CPT	dominante	0.488; 0.551	1.43 (1.07-1.91)	1.5×10^{-2}	0.476; 0.472	0.93 (0.75-1.16)	0.522
rs4939827	<i>SMAD7</i>	CPT	recesivo	0.461; 0.416	0.68 (0.50-0.94)	1.9×10^{-2}	0.438; 0.478	1.32 (1.03-1.68)	0.027
rs2066807	<i>STAT2</i>	CPT	dominante	0.033; 0.054	1.69 (1.06-2.69)	2.6×10^{-2}	0.040; 0.039	0.98 (0.68-1.42)	0.928
rs2284734	<i>TSHR</i>	CPT clásico	recesivo	0.290; 0.388	2.64 (1.69-4.13)	1.8×10^{-5}	0.304; 0.317	1.42 (0.99-2.03)	0.058
rs1053266	<i>CCDC6</i>	CPT clásico	recesivo	0.494; 0.543	1.37 (0.96-1.97)	8.7×10^{-2}	0.471; 0.502	1.24 (0.96-1.60)	0.103
rs2687834	<i>TG</i>	CPTvf	recesivo	0.447; 0.565	2.28 (1.50-3.46)	1.1×10^{-4}	0.491; 0.477	0.74 (0.46-1.17)	0.195
rs6179	<i>GHR</i>	CFT	dominante	0.356; 0.232	0.46 (0.27-0.80)	5.4×10^{-3}	0.299; 0.286	0.87 (0.55-1.37)	0.543
rs13099828	<i>PPARG</i>	CPTvf+CFT	dominante	0.186; 0.256	1.71 (1.22-2.38)	1.6×10^{-3}	0.168; 0.189	1.01 (0.71-1.46)	0.936
rs25882	<i>CSF2</i>	agresivos	dominante	0.185; 0.129	0.69 (0.48-0.99)	4.7×10^{-2}	0.194; 0.214	1.08 (0.74-1.57)	0.689

OR obtenidos al comparar casos frente a controles considerando el alelo polimórfico como el alelo de riesgo. ^a Resultados ajustados por las covariables edad, género y origen. Abreviaturas: MAF= *minor allele frequency*; OR= *odds ratio*; IC 95%= intervalo de confianza al 95 %; CPT=carcinoma papilar de tiroides; CPTvf= variante folicular del CPT; CFT= carcinoma folicular de tiroides.

4.2.2. Resultados en Carcinoma Papilar de Tiroides clásico (CPTc)

El subtipo clásico del CPT estaba bien representado en nuestra serie de validación (n=582), y nos permitió disponer de la potencia teórica suficiente para detectar asociaciones.

Para el tagSNP de *TSHR*, obtuvimos resultados marginalmente significativos ($P=0.058$), los cuales sugieren que ésta u otra variante potencialmente funcional podrían jugar un papel en la susceptibilidad de desarrollar CPTc. De forma similar, aunque menos consistente, ya que desaparecía al ajustar los resultados, obtuvimos una tendencia (P no ajustada=0.070) para el otro polimorfismo estudiado en CPTc: el cambio de aminoácido en *CCDC6*, siendo éste otro buen candidato a actuar como GBP.

4.2.3. Resultados en la variante folicular del Carcinoma Papilar de Tiroides, (CPTvf)

Los resultados de validación en CPTvf estuvieron condicionados por la pobre caracterización de este subtipo en nuestra serie de validación, en la que disponíamos sólo de 118 casos diagnosticados con seguridad para este subtipo. Independientemente del bajo tamaño muestral, no obtuvimos ni siquiera una tendencia para el tagSNP en el gen *TG*, considerado en este subgrupo.

4.2.4. Resultados en Carcinoma Folicular de Tiroides (CFT)

La serie de CFT, teniendo en cuenta su poca prevalencia, estuvo bien representada en la serie de validación (n=104). Sin embargo, no se obtuvieron resultados significativos para la variante exónica en *GHR* estudiada.

4.2.5. Resultados en carcinomas con patrón de crecimiento folicular (CPTvf y CFT)

A continuación, y de forma similar a la serie 1, consideramos juntos los carcinomas con patrón de crecimiento folicular, es decir, CPTvf y CFT, que ascendió a

222 casos. El estudio de la distribución de frecuencias del polimorfismo en *PPARG* en estos casos frente a controles produjo resultados no significativos.

4.2.6. Resultados en carcinomas agresivos

Por último, estudiamos la variante de cambio de aminoácido en *CSF2* en aquéllos carcinomas que hubiesen presentado algún tipo de metástasis (n=152). Sin embargo, los resultados obtenidos en la serie 1 no se replicaron en las series de validación, descartando este polimorfismo como un factor de protección.

4.2.7. Resultados atendiendo a características clínicas del paciente

Debido a la ausencia de validación para *CSF2* en carcinomas agresivos, nos planteamos ampliar el estudio, considerando exclusivamente los casos. Para ello, nos preguntamos si alguna de las variantes genéticas consideradas en los estudios caso-control podría tener un efecto sobre ciertas características de los tumores. Así, no sólo estudiamos la agresividad (entendida como extensión extratiroidea), sino también otras características de los tumores, como la presencia de bilateralidad, multifocalidad o la edad de aparición de la enfermedad. Para ello, realizamos un estudio caso-caso, considerando de forma conjunta los pacientes de las series 1, 2 y 3 (n=1629).

En primer lugar, comparamos aquellos casos que presentasen múltiples focos tumorales y/o bilateralidad (n=200) frente a aquéllos que no (n=1429). Observamos que en el primer grupo, los genotipos rs4939827CC de *SMAD7* y rs721992GT+TT de *CCDC6* estaban significativamente infrarrepresentados ($P=0.030$ y $P=0.009$, respectivamente).

Por otra parte, dividimos los casos entre aquéllos que hubiesen presentado algún tipo de metástasis, tanto a nódulos linfáticos como a distancia (n=369), y los que no presentaron extensión extratiroidea (n=672). En el grupo de los carcinomas más agresivos se observó una menor representación del genotipo rs2703488CC del gen *KIT* ($P=0.026$). Sin embargo, no se encontró significación para la variante rs25882 de *CSF2*, confirmando la falta de validación observada en las series 2+3 del estudio caso-control.

Finalmente, tratamos de averiguar si algún polimorfismo podría influir en la edad de aparición de la enfermedad. Para ello, establecimos un punto de corte en los 40 años al diagnóstico, dividiendo nuestra serie entre casos tempranos (n=600) y casos tardíos (n=880). En el grupo de pacientes diagnosticados con 40 años o menos estaban significativamente sobrerrepresentados los genotipos rs25882TC+CC de *CSF2* ($P=0.015$), rs6179GA+AA de *GHR* ($P=0.008$) y rs4422036AA de *MED13* ($P=0.045$).

5. Resultados del estudio de interacción gen-gen

Debido a la ausencia de validación para la mayoría de nuestras variantes, y teniendo en cuenta el carácter complejo de esta enfermedad, nos planteamos un estudio de epistasia o interacción gen-gen, que pudiese explicar, al menos en parte, la susceptibilidad genética a desarrollar estos carcinomas.

Para la detección de interacciones, empleamos una aproximación similar a la realizada para SNPs individuales, utilizando la serie 1 como fase de descubrimiento, y las series 2+3, como validación. Como al considerar parejas de SNPs en lugar de variantes individuales, las combinaciones de genotipos aumenta exponencialmente (de 3 a 9 posibles), restringimos este análisis a la comparación CPT *versus* controles, en la que disponemos de un tamaño muestral mayor.

En primer lugar, utilizamos el programa MB-MDR para detectar parejas de SNPs candidatas a conferir una mayor susceptibilidad al desarrollo de CPT. De esta forma, detectamos 25 interacciones candidatas, que implicaban a 17 SNPs, localizados en 16 genes distintos, de los que sólo la mitad habían sido identificados en el estudio de efectos individuales. Los resultados están resumidos en la figura R17, en la que también puede apreciarse cómo aparecieron nuevas variantes y genes candidatos.

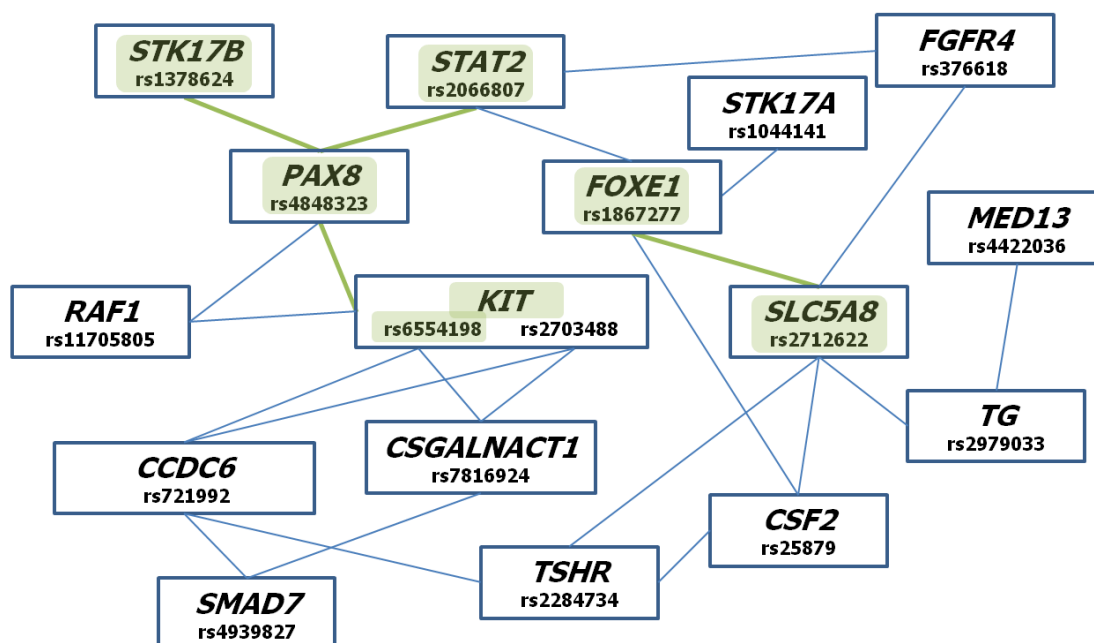


Figura R17. Red de interacciones gen-gen identificadas en la serie 1. Las interacciones están ajustadas por género, edad y por el efecto individual de cada SNP. Se realizaron 10000 permutaciones para calcular la P , y se estableció un umbral de $P < 0.10$. En verde, se resaltan las interacciones, genes y SNPs que posteriormente se replicaron en las series 2+3.

Dado el elevado número de comparaciones realizadas (269,745 pares posibles), resulta esencial replicar los datos de interacción. Por ello, utilizamos las series 2+3 para validar los resultados obtenidos. Para ello, con el mismo programa, realizamos el análisis de las 25 interacciones candidatas, evaluando así si las mismas combinaciones de genotipos de riesgo o protección se mantenían en este nuevo grupo. Los resultados, resumidos en la tabla R17, mostraron que cuatro de ellas se mantenían estadísticamente significativas en esta segunda fase (figura R18, tabla R17). Es muy importante destacar que dichas parejas de genes actuaban además conforme a los mismos modelos de la serie 1.

Como puede observarse en la tabla R17, tres de las cuatro interacciones validadas implicaban al gen *PAX8* (*paired box 8*), un factor de transcripción relacionado con la expresión de genes específicos de tiroides. Además, se observó un efecto epistático para el SNP rs1867277 de *FOXE1*, además del importante efecto individual descrito anteriormente.

Tabla R17. Resultados de las cuatro interacciones gen-gen validadas.

SNP, gen #1	Función #1	SNP, gen #2	Función #2	#1+2 ^a	Efecto	P_1 ^b	P_{2+3} ^b
rs4848323 <i>PAX8</i>	tagSNP, intrón 2	rs6554198 <i>KIT</i>	tagSNP, 5' <i>ups</i>	CC+GG	Riesgo	0.060	0.007
rs4848323 <i>PAX8</i>	tagSNP, intrón 2	rs1378624 <i>STK17B</i>	tagSNP, 5' <i>ups</i>	CT+GG	Protección	0.036	0.041
rs1867277 <i>FOXE1</i>	TFBS, 5' UTR	rs2712622 <i>SLC5A8</i>	tagSNP, intrón 1	GG+CC	Riesgo	0.056	0.014
rs4848323 <i>PAX8</i>	tagSNP, intrón 2	rs2066807 <i>STAT2</i>	M594I, exón 20	CC+CG	Protección	0.036	0.042

^a Las combinaciones de genotipos de riesgo o protección están representadas en la figura R18.

^b Los valores de P para las series 1 (P_1) y 2+3 (P_{2+3}) están ajustados por género, edad y efecto individual de cada SNP, y corregidos por 10000 permutaciones.

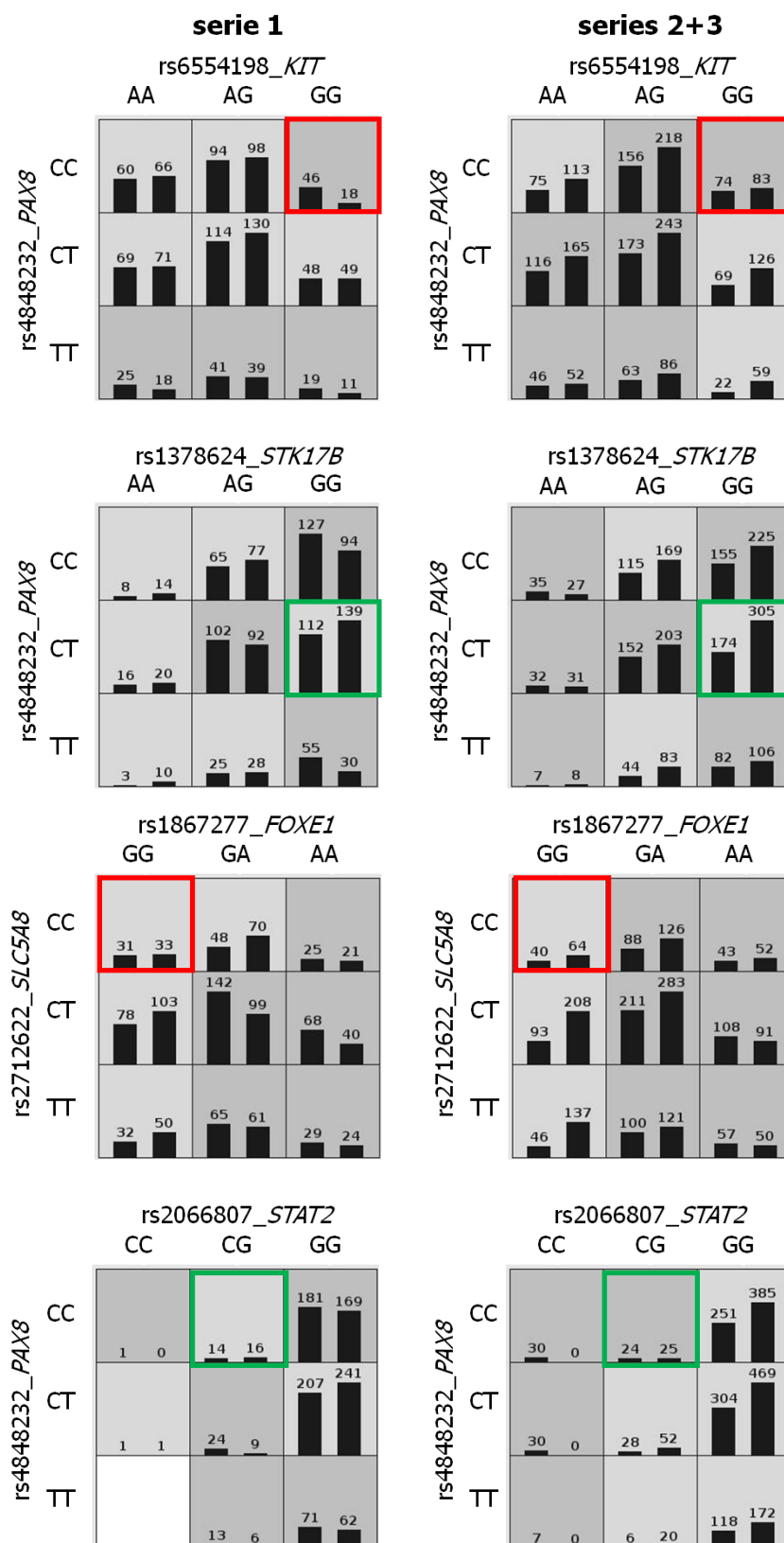
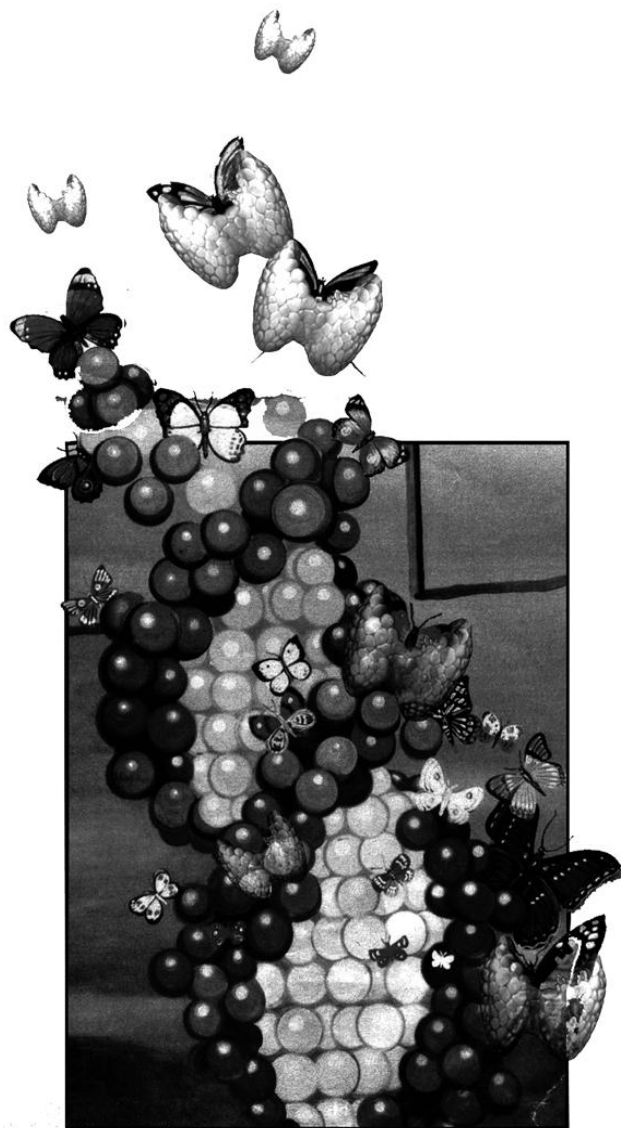


Figura R18. Resultados de epistasis para las cuatro interacciones validadas. Los recuadros rojos indican combinaciones de riesgo, y los verdes, protección. Notar que estas representaciones, obtenidas con MDR, no están corregidas por el efecto individual de cada SNP. Para más detalles, ver tabla R17.

DISCUSIÓN



La identificación de factores genéticos de riesgo a desarrollar carcinomas de tiroides de estirpe folicular es una tarea ardua debido a la heterogeneidad y a la frecuencia relativamente baja con la que aparecen estas neoplasias. Esta enfermedad está ampliamente considerada como una entidad compleja, en la que variantes genéticas comunes, localizadas en genes de baja penetrancia (GBP) interactuarían entre sí y con factores ambientales, determinando la susceptibilidad individual (Adjadj, et al., 2009; Landa and Robledo, 2011; Sturgis and Li, 2009). En este contexto, los análisis de ligamiento no tienen la potencia suficiente para detectar las contribuciones genéticas moderadas que confieren los GBP, por lo que la estrategia más adecuada para abordar la identificación de estos factores genéticos de riesgo son los estudios de asociación (Houlston and Peto, 2004; Milne and Benitez, 2008).

Hasta la fecha, muchos de los estudios de asociación publicados en carcinomas de tiroides de estirpe folicular tienen una potencia limitada, consecuencia de tamaños muestrales moderados. Así, se han descrito variantes asociadas al desarrollo de estos carcinomas en los genes *RET* (Ho, et al., 2005; Lesueur, et al., 2002), *ATM* (Akulevich, et al., 2009), *CYP1A1* (Bufalo, et al., 2006; Siraj, et al., 2008b), *VEGFA* (Hsiao, et al., 2007), *CYP2D6* (Lemos, et al., 2007), *NAT2* (Hernandez, et al., 2008), *RAD52* (Siraj, et al., 2008a), *GSTT1* (Lemos, et al., 2008; Siraj, et al., 2008b), *VDR* (Penna-Martinez, et al., 2009), *CHEK2* (Cybulski, et al., 2004) y *GSTM1* (Lemos, et al., 2008; Siraj, et al., 2008b), si bien todos estos GBP potenciales necesitarían ser validados.

Además del tamaño muestral, probablemente uno de los problemas relacionados con la falta de validación de unas series a otras esté relacionado con haber considerado esta enfermedad como un todo, y no haber distinguido entre variantes histológicas u otro tipo de características del tumor. En nuestro estudio, tratando de tener en cuenta estas consideraciones, hemos recopilado una serie de muestras que supone el grupo más amplio de este tipo de pacientes estudiado hasta la fecha. Además, hemos considerado, en la medida de lo posible, la heterogeneidad de estos tumores, subestratificando el grupo de pacientes en función de sus características específicas, tratando así de evitar el ruido de fondo que existe cuando se mezclan casos con características diferentes. A pesar de las limitaciones que suponen el carácter complejo y la heterogeneidad de esta enfermedad, hemos sido capaces de identificar varios factores genéticos de riesgo a desarrollar cáncer de tiroides. Entre ellos, destacan la asociación del polimorfismo rs34330 del gen *CDKN1B* en la susceptibilidad de desarrollar CPTvf, y la de la variante rs1867277 de *FOXE1* en el

desarrollo de CPT. En los siguientes apartados se comenta el papel de estos GBP, entre otros, con especial atención a su efecto funcional, tratando así de desvelar los mecanismos moleculares implicados en la susceptibilidad a estas neoplasias.

Por otra parte, dado el carácter complejo de esta enfermedad, parece poco probable que esté determinada por una o unas pocas variantes genéticas, sino más bien por la combinación de muchas. Sin embargo, hasta ahora, la susceptibilidad al carcinoma de tiroides de células foliculares se ha tratado de una forma bastante *naïve*. Por eso, nuestro trabajo trata de dar una visión más integradora de la enfermedad, identificando, a través de un estudio de epistasia, interacciones entre varios genes. A pesar de las limitaciones técnicas que estos análisis conllevan, pensamos que deberían ser generalizados de aquí en adelante para intentar descifrar las bases genéticas subyacentes a esta enfermedad.

1. La variante rs34330 (-79C>T) en el gen *CDKN1B* (p27) confiere riesgo a desarrollar CPTvf y altera los niveles de ARNm

Nuestro estudio es el primero que ha investigado el papel de variantes polimórficas en el *locus* de *CDKN1B* (p27) en la susceptibilidad a desarrollar cáncer de tiroides. El análisis de regresión logística nos permitió identificar el SNP rs34330 (c.-79C>T), localizado en el promotor del gen *CDKN1B*, como significativamente asociado al desarrollo de la variante folicular del carcinoma papilar de tiroides (CPTvf), ajustándose a un modelo recesivo (OR=2.12; $P=0.023$). La baja correlación ($r^2 \rightarrow 0$) con los SNPs cercanos a este polimorfismo, junto a los resultados no significativos para rs2066827 y rs36228499, minimiza la posibilidad de que otras variantes sean las causales.

El polimorfismo rs34330 se ha asociado previamente a otras neoplasias. Chang y colaboradores describieron una asociación de este SNP en cáncer de próstata hereditario (Chang, et al., 2004). En dicho trabajo, el alelo C de la variante era transmitido de los padres a su descendencia afecta en una proporción superior a la esperada. También sugerían que el polimorfismo era capaz de perturbar un sitio de metilación (isla CpG), hipótesis que no ha sido demostrada hasta ahora. Merece la pena destacar que un estudio de asociación en dos pasos, que incluyó más de 2200 pacientes y más de 2200 controles, identificó la misma variante asociada significativamente al desarrollo de cáncer de mama en población británica (Driver, et al., 2008). Estos autores describieron el mismo genotipo de riesgo (TT) identificado en nuestro trabajo, ajustándose también a un modelo recesivo (OR=1.22; $P=0.013$). En otro trabajo reciente, esta variante se relacionó con riesgo a cáncer de pulmón (OR=1.27; IC 95%=1.01-1.60) (Wang, et al., 2007a), que sería especialmente alto para individuos “muy fumadores”, definidos como aquéllos que hubiesen fumado al menos 37.5 paquetes de tabaco (OR=6.24; IC 95%=1.67–23.34). Propusieron que, debido a la localización de este SNP, su efecto funcional podría estar asociado a la reducción en la producción de la proteína p27. De hecho, mencionan que el SNP está localizado en la región 5' UTR, donde Millard y colaboradores (Millard, et al., 2000) habían descrito la existencia de un “elemento rico en uracilos”, que permitiría la unión a varios factores (HuR, hnRNP C1 y C2), involucrados en la estabilidad, procesamiento y traducción del ARNm. Sin embargo, el polimorfismo rs34330 se localiza a 27 nucleótidos *upstream* de la secuencia consenso del “elemento rico en uracilos” definido por Millard. Este hecho, junto con algunas predicciones *in silico* contradictorias que

obtuvimos para los factores de unión mencionados, nos llevó a la conclusión de que éste no era el mecanismo subyacente al efecto del SNP.

Nuestra hipótesis fue que el polimorfismo rs34330 podría alterar la transcripción de *CDKN1B*. De hecho, dos herramientas bioinformáticas independientes (*CONSITE* y *TFSEARCH*), integradas en la base de datos F-SNP, predecían que el rs34330 estaba localizado en un sitio importante para la unión de factores de transcripción. Además, ambas herramientas predecían una afinidad y unión diferenciales para los factores de transcripción humanos AP-2 α y SP1 a la secuencia con la variante genética. La unión estaría fuertemente condicionada por la presencia del polimorfismo, especialmente para el primer factor de transcripción.

Las predicciones anteriores estaban respaldadas por el mecanismo de regulación descrito para ciertos genes, que contienen secuencias solapadas con sitios de unión para AP-2 y SP1 (Hilger-Eversheim, et al., 2000). En estos casos, la transcripción está regulada por la proporción SP1/AP-2. Esto es de especial interés, teniendo en cuenta que se sabe que AP-2 α es capaz de unirse a *CDKN1A* (p21^{cip1}) (Zeng, et al., 1997), gen que codifica un inhibidor del ciclo celular que pertenece a la misma familia que *CDKN1B* (p27^{kip1}). Los tres experimentos realizados en nuestro estudio (figuras R1-3): 1. ensayos de luciferasa con el promotor polimórfico y *wildtype* en células HeLa; 2. actividad del promotor de *CDKN1B* en células PCC13; y 3. cuantificación de niveles de ARN mensajero de *CDKN1B* en voluntarios sanos; proporcionan evidencias suficientes para concluir que la variante rs34330 (c.-79C>T) influye significativamente en las tasas de transcripción de *CDKN1B*. Por tanto, debe ser considerado como un mecanismo alternativo de regulación negativa en cáncer de tiroides, independientemente de los factores de transcripción precisos que estén involucrados (figura D1). Este hallazgo es de especial relevancia porque, aunque p27 ha sido ampliamente estudiado en tumores (Chu, et al., 2008; Slingerland and Pagano, 2000), y específicamente en cáncer de tiroides (Erickson, et al., 1998; Tallini, et al., 1999; Vitagliano, et al., 2004), hasta ahora solo se habían descrito alteraciones en la traducción, degradación y deslocalización de p27, siendo ésta la primera vez que se sugiere una tasa de transcripción de *CDKN1B* dependiente de una variante genética polimórfica localizada en su promotor, que debería ser estudiado con mayor profundidad en cáncer.

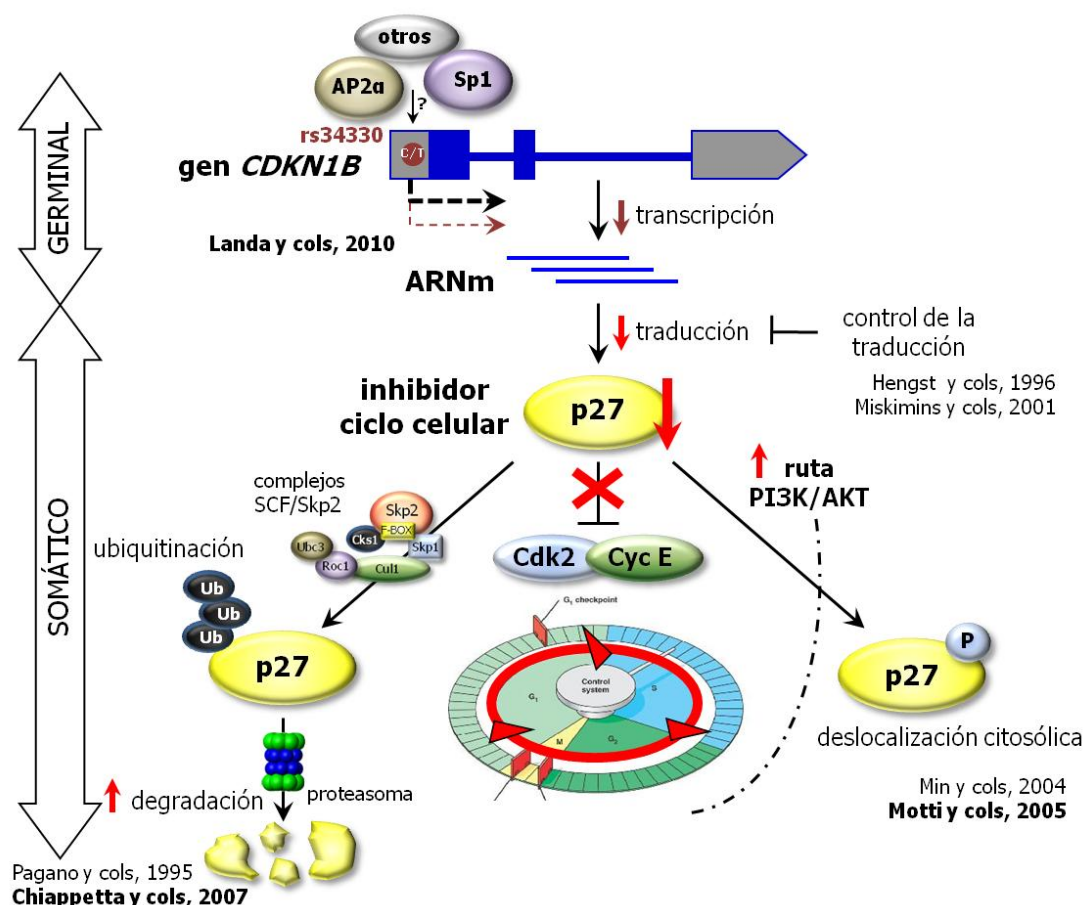


Figura D1. Mecanismos de regulación negativa de p27. La disponibilidad de p27 en tumores, y por tanto, su capacidad para inhibir la progresión del ciclo celular, está reducida mediante varios mecanismos: inhibición de la traducción, aumento en su degradación, o deslocalización al citosol. Nosotros proponemos otro mecanismo, a nivel germinal, de disminución de p27: una reducción en su transcripción, condicionada por la presencia de la variante de riesgo rs34330 (c.-79C>T) del gen *CDKN1B*. En negrita, se muestran las referencias a los trabajos que han descrito esos mecanismos específicamente en cáncer de tiroides.

Resulta especialmente atractivo que una variante polimórfica se asocie específicamente con un subtipo histológico tumoral, en este caso la variante folicular del carcinoma papilar de tiroides (CPTvf). Estos resultados apuntan a que la subestratificación del fenotipo será un punto esencial en la identificación de genes de baja penetrancia, y probablemente en la validación en series independientes. Es decir, esta observación proporciona una base para la heterogeneidad descrita en esta neoplasia, que aglutina más de 12 subtipos histológicos. Una posible explicación biológica podría ser que las alteraciones genéticas características de CPTvf serían similares a las descritas en CFT y, por tanto, relacionadas con la ruta metabólica de PI3K/Akt. A este respecto, hay varias evidencias que sugieren que la desregulación de

la vía PI3K/Akt juega un papel importante en la patogénesis de tumores tiroideos, especialmente en CFT. (Hou, et al., 2007; Liu, et al., 2008; Riesco-Eizaguirre and Santisteban, 2007; Wang, et al., 2007b). Dado que p27 pertenece a esta ruta, una alteración en la tasa de transcripción de *CDKN1B*, causada por la variante rs34330, podría ser de especial importancia con estos tumores.

Respecto a los dos polimorfismos de *CDKN1B* para los que no encontramos una asociación significativa con cáncer de tiroides, no hay estudios sobre la influencia del rs36228499 (-838C>A) en cáncer. Sólo un trabajo (Spurdle, et al., 2008) muestra una evidencia indirecta de que este SNP, representado por el tagSNP vecino rs3759216, no modifica el riesgo a desarrollar cáncer de mama entre los portadores de mutación en *BRCA1* o *BRCA2*. Por el contrario, hay muchos estudios que describen asociación significativa entre el SNP codificante rs2066827 (p.V109G) y el desarrollo de cáncer. Entre ellos, merece la pena mencionar los relacionados con carcinoma avanzado de próstata (Kibel, et al., 2003), carcinoma oral de células escamosas (Li, et al., 2004), cáncer invasivo del epitelio de ovario (Gayther, et al., 2007), tumores de mama de grado alto (Tigli, et al., 2005) y metástasis a nódulos linfáticos en cáncer de mama (Naidu, et al., 2007). Sin embargo, nuestros datos, basados en una amplia serie española, sugieren que este cambio de aminoácido en p27 no se asocia al menos a ningún subtipo de cáncer de tiroides.

2. La variante rs1867277 en el gen *FOXE1* confiere susceptibilidad a desarrollar cáncer de tiroides mediante el reclutamiento de los factores de transcripción USF1/USF2

El estudio de asociación en dos pasos, utilizando una aproximación de genes candidatos, identificó a *FOXE1* como un gen de baja penetrancia (GBP) relacionado con al desarrollo de carcinoma papilar de tiroides (CPT). El análisis de regresión nos permitió identificar varios SNPs en el *locus* de *FOXE1* como variantes altamente significativas que confieren riesgo a desarrollar CPT. El análisis de haplotipos de *FOXE1*, seguido de una búsqueda exhaustiva de una variante potencialmente funcional, permitió identificar y seleccionar el polimorfismo de región promotora (rs1867277, c.-283G>A) para posteriores estudios. Los resultados de asociación se validaron en dos poblaciones independientes (series 2 y 3) (tablas R15 y R16), alcanzando el OR combinado por alelo de 1.45 (IC 95 %=1.30-1.61; $P=4.7 \times 10^{-12}$).

El factor de transcripción FoxE1 pertenece a la familia de factores de transcripción *forkhead*. Estos factores comparten una estructura de unión al ADN, conocido como "dominio de hélice alada" (*winged helix domain*), capaz de interactuar con los nucleosomas, creando un dominio local expuesto necesario para la acción de otros factores de transcripción. Dicha propiedad define a FoxE1 como un factor de transcripción pionero (Cuesta, et al., 2007), cuya acción es esencial para el desarrollo, diferenciación y respuesta a hormonas de la glándula tiroidea (De Felice and Di Lauro, 2004). Por tanto, el control de su expresión debe estar regulado de forma exquisita. Sin embargo, existen pocos datos disponibles en relación a los factores de transcripción involucrados en el control de la expresión de *FOXE1* en el tiroides, aunque el hecho de que la hormona TSH controle la expresión de *FOXE1* a través del AMP cíclico y el Ca^{2+} (Aza-Blanc, et al., 1993; Ortiz, et al., 1997) sugiere que los factores de unión a elementos CRE ("*calcium response element*") y a Ca^{2+} podrían jugar un papel. De hecho, el análisis de la secuencia del promotor de *FOXE1*, donde está localizado el rs1867277 (-c.283G>A), revelaron la existencia de un sitio DRE, similar al previamente descrito por D'Andrea y cols (D'Andrea, et al., 2005), y que compartía una alta similitud con el sitio consenso DRE en el gen de la prodinorfina (Jacobson, et al., 2006). Sin embargo, los ensayos funcionales mostraron que el factor dependiente de calcio DREAM se unía con la misma afinidad a ambos alelos del rs1867277 (figura R13), sugiriendo otro mecanismo de acción. De hecho, finalmente pudo demostrarse la formación de un complejo ADN-proteína exclusivamente entre el

alelo rs1867277A y los factores USF1/USF2. Estas proteínas tienen una expresión ubicua y pertenecen a la familia de factores de transcripción hélice-vuelta-hélice (*helix-loop-helix*, *HLH*) de cremallera de leucinas. Dichos factores comparten una región C-terminal altamente conservada, responsable de su dimerización y de la unión al ADN, que reconoce la secuencia canónica CACGTG de la *E-box* (Rada-Iglesias, et al., 2008; Viollet, et al., 1996). La implicación de los factores USF en la regulación de *FOXE1* se confirmó mediante los ensayos de transfección. Estos factores, que actúan sobre todo como heterodímeros, indujeron aumentos significativos en la actividad transcripcional de *FOXE1* en presencia del alelo rs1867277A. Estos datos concuerdan, por una parte, con el papel de los factores USF como reguladores transcripcionales positivos de sus genes diana (Runkel, et al., 1991), y por otra, con la acción predominante de los heterodímeros USF1/USF2 en comparación con los homodímeros (Viollet, et al., 1996).

Además, se han descrito interacciones entre los dominios N-terminal de los dímeros USF y factores de transcripción específicos de célula, lo que llevaría a una regulación transcripcional cooperativa (Corre and Galibert, 2005). Es interesante señalar que identificamos un sitio CRE potencial localizado cerca de la variante rs1867277, que podría actuar como diana de proteínas relacionadas con CRE (CREB y α CREM). Los ensayos de transfección transitoria demostraron que CREB, y aún más α CREM, activan el promotor de *FOXE1*, mientras que DREAM reduce significativamente la inducción transcripcional dependiente de α CREM. Estos datos sugieren la existencia de una competición directa por la unión al promotor de *FOXE1* entre los factores de unión a CRE y DREAM, lo que determinaría en última instancia el estatus transcripcional del gen *FOXE1* (figura D2). De hecho, se ha descrito que el Ca^{2+} y el AMPc modulan una red reguladora que relaciona a las proteínas de unión a CRE con el factor DREAM (Ledo, et al., 2000; Ledo, et al., 2002). En este sentido, dichos factores de transcripción podrían regular la expresión de *FOXE1* en respuesta a TSH en situación fisiológica.

Dado que tanto las proteínas CRE como los factores USF están asociados con un incremento en los niveles de expresión de *FOXE1* a través de sitios de unión adyacentes, surge la cuestión de si estos factores, además, cooperan para regular la transcripción de *FOXE1*. Nuestros resultados mostraron la ausencia de un efecto sinérgico entre las proteínas α CREM y USF1/2 sobre la transcripción de *FOXE1*, lo cual está de acuerdo con lo descrito en otros trabajos (Mestre, et al., 2001; Scholtz, et al.,

1996). Por tanto, queda por elucidar qué otras proteínas, si las hay, y de qué manera interactúan con los factores USF para modular la expresión de *FOXE1* (figura D2).

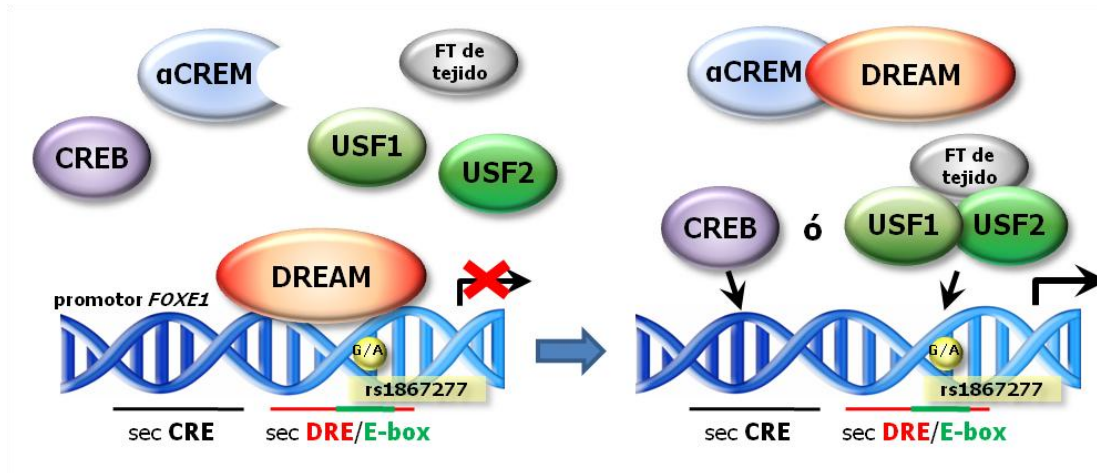


Figura D2. Mecanismo de regulación propuesto para el promotor de *FOXE1*, según nuestros resultados y lo descrito en otras publicaciones. DREAM actúa como represor de la activación mediada por factores USF y factores CRE. La proteína α CREM es capaz de secuestrar el inhibidor DREAM, facilitando la transcripción de *FOXE1* y permitiendo la activación del promotor por CREB. Los factores USF se unen a su secuencia consenso sólo en presencia del alelo de riesgo rs1867277A, activando la transcripción de *FOXE1* por un mecanismo independiente de los factores CRE, y que requiere posiblemente de otros factores de transcripción específicos de tejido. Globalmente, la presencia del SNP produce un cambio en el balance de la regulación de la expresión de *FOXE1*, lo que produciría un aumento en la susceptibilidad a desarrollar carcinoma papilar de tiroides.

Finalmente, y considerando la unión específica de los factores USF al alelo de riesgo rs1867277A, sería esperable un aumento en la expresión de *FOXE1* en tumores de células foliculares de tiroides comparados con tirocitos normales. Existen muy pocos datos sobre el estatus de *FOXE1* en cáncer de tiroides, aunque Sequeira y colaboradores demostraron que la expresión de *FOXE1* aumentaba en paralelo con el proceso de dediferenciación de los carcinomas de tiroides (Sequeira, et al., 2001). Por tanto, es necesario entender de qué manera FoxE1, un factor de transcripción tiroideo implicado en el mantenimiento del fenotipo diferenciado del tiroides adulto, podría estar implicado en la adquisición de un estatus maligno. Una posible explicación puede obtenerse de los resultados en modelos de ratón *knockout* para *FOXE1*. Durante el desarrollo embrionario, las células precursoras tiroideas necesitan la transcripción de *FOXE1* para permitir inicialmente su propia migración desde el primordio tiroideo, y posteriormente para constituir el órgano endocrino funcional (De Felice, et al., 1998;

Parlato, et al., 2004). Considerando estos datos, la hipótesis que planteamos es que el aumento de expresión de *FOXE1* en carcinomas de tiroides podría estar relacionado con una ventaja en la motilidad de las células tiroideas malignas, lo cual estaría favorecido por la presencia del alelo de riesgo rs1867277A. Aunque se necesitan estudios específicos sobre FoxE1 para entender su papel en la migración e invasión de células tumorales, varios estudios han confirmado el papel en tumorigénesis de los factores de transcripción de la familia *forkhead box* (Fox). A este respecto, varios genes que codifican factores *forkhead*, junto a otros genes, han sido recientemente identificados como una firma molecular de la transición epitelio-mesénquima en cáncer de colon (Joyce, et al., 2009). Además, la sobreexpresión de factores Fox se ha descrito en varias neoplasias (Barrans, et al., 2004; Curiel, et al., 2004; Heselmeyer, et al., 1997; Mani, et al., 2007; Rodriguez, et al., 2005; Singh, et al., 2001; Spirin, et al., 1996; Wlodarska, et al., 2005). Esto abre un futuro interesante para comprender el papel de *FOXE1* en la capacidad de migración e invasión de las células tumorales tiroideas.

En resumen, nuestro estudio de asociación, combinado con la evaluación funcional, nos permitió identificar una variante funcional en el promotor de *FOXE1*, y proponer un mecanismo por el cual esta variante funcional actúa como un factor de riesgo específico en la susceptibilidad a desarrollar CPT. Este hallazgo destaca de nuevo la importancia de la selección de los pacientes, mediante una clasificación clínica homogénea y rigurosa, que permita un análisis estratificado de fenotipos. También es importante recordar que la naturaleza compleja de cada *locus* podría operar a través de variantes o mecanismos adicionales.

Por último, hay que destacar que la identificación de *FOXE1* como un gen de baja penetrancia relacionado con el desarrollo de CPT representa, probablemente, el ejemplo más consistente, más validado y mejor caracterizado funcionalmente de todos los publicados hasta ahora en estas neoplasias. Así, en los últimos dos años, además de nuestro estudio de genes candidatos (Landa, et al., 2009), los dos únicos estudios de asociación del genoma completo (GWAS) publicados en cáncer de tiroides, han obtenido sus resultados más significativos para variantes situadas en el *locus* de *FOXE1* (Gudmundsson, et al., 2009; Takahashi, et al., 2010). En el primer GWAS realizado en estos carcinomas, se asoció significativamente el tagSNP intergénico rs965513, localizado en la región 9q22.33, a 57 kb *upstream* de *FOXE1*, al desarrollo de CPT. Los autores obtuvieron un OR=1.75 (IC 95 %=1.59-1.94; $P=1.7 \times 10^{-27}$), basado en un

grupo de 962 pacientes (CPT+CFT), y correlacionaron el alelo de riesgo con bajas concentraciones en sangre de TSH y T4, y niveles altos de T3 (Gudmundsson, et al., 2009). En paralelo, nuestro estudio, basado inicialmente en 984 CPT, relacionó la variante funcional rs1867277 (c.-283G>A), localizada en la región 5'UTR de *FOXE1*, con un aumento significativo en el riesgo a desarrollar la enfermedad (OR=1.49; IC 95 %=1.30–1.70; $P=5.9 \times 10^{-9}$), proponiendo además, como se ha comentado, un mecanismo molecular subyacente, basado en el reclutamiento diferencial de los factores de transcripción USF1/USF2 (Landa, et al., 2009). Finalmente, un tercer estudio, que incluyó 660 CPT asociados a radiación, procedentes de las áreas afectadas por el accidente de Chernobyl, extendió y validó los resultados anteriores, tanto para la variante rs965513 (OR=1.65; IC 95 %=1.43-1.91; $P=4.8 \times 10^{-12}$), como para el rs1867277 (OR=1.48; IC 95 %=1.27-1.71; $P=4.5 \times 10^{-7}$). Los importantes tamaños muestrales considerados, los OR y modelos de herencia similares obtenidos en estos tres estudios, junto con la evaluación funcional de las variantes de *FOXE1*, destacan la consistencia de este *locus* como un gen de baja penetrancia en CPT. La relevancia de nuestro estudio respecto a los otros dos está en que aborda una de las mayores limitaciones de los GWAS: la enorme dificultad para establecer un vínculo entre un tagSNP intergénico y una variante concreta con un papel causal y una explicación biológica definidos (Frazer, et al., 2009). Además, supone un buen ejemplo de cómo un *locus* biológicamente relevante *a priori*, ha sido ampliamente identificado como un GBP. Queda por elucidar, comparando nuestros resultados con los obtenidos por otros grupos (tabla S2), si hay una o más variantes funcionales que condicionan la función de *FOXE1*.

3. Otros genes podrían determinar la susceptibilidad a desarrollar carcinomas de tiroides de estirpe folicular

En este trabajo, además de los genes *CDKN1B* y *FOXE1*, hemos identificado otros candidatos a actuar como GBP en carcinomas de tiroides de estirpe folicular. Si bien ninguna de las variantes elegidas en estos genes fue validada en la segunda fase del estudio, parece prematuro descartarlos sin intentar antes validarlos con series más amplias y mejor caracterizadas.

3.1. *TSHR*: gen del receptor de la tirotropina

Entre las señales muy significativas de asociación detectadas en la serie 1, destacamos las variantes de *TSHR*, que codifica para el receptor de la tirotropina u hormona estimuladora del tiroides (TSH), y lo proponemos para ser estudiado en mayor detalle y determinar así su posible papel en el riesgo a desarrollar CPT. Como se muestra en la figura D3, las variantes de *TSHR* se localizan en una región con un alto LD, que complica la identificación de la posible variante funcional. Además, se debería tener en cuenta el *splicing* alternativo que determina la proporción de las dos isoformas de *TSHR* descritas. Una de estas isoformas carece de toda la región intracelular y transmembrana del receptor, y por tanto no es funcional. Una variante que afectase la proporción entre la forma activa o la no funcional, explicaría la señal de asociación identificada entre el grupo de pacientes con CPT, y debería ser estudiado en detalle.

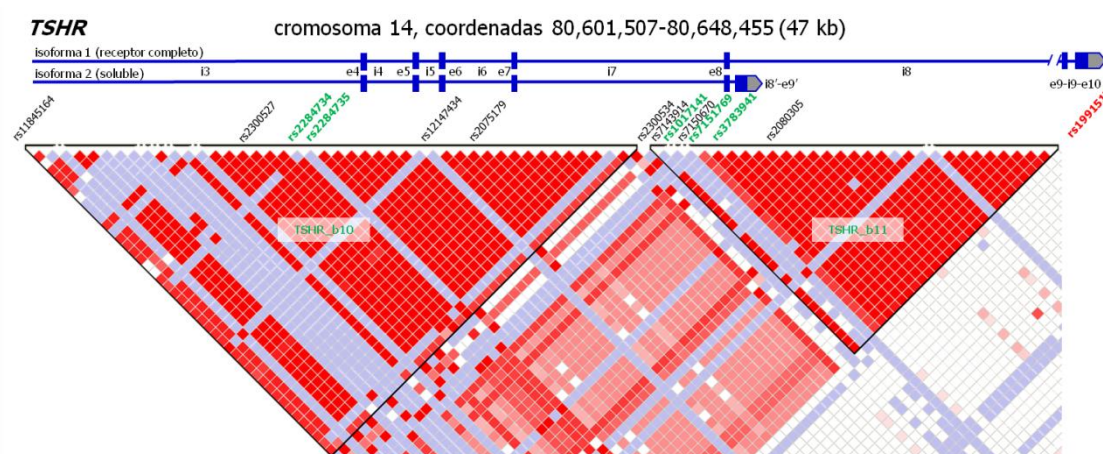


Figura D3. Detalle del *locus* de *TSHR*. Se muestra la localización de las dos isoformas codificantes descritas para este gen, su patrón de disequilibrio de ligamiento, y los polimorfismos estudiados. En verde, se resaltan las asociaciones muy significativas en CPT clásico, tanto para SNPs individuales, como para haplotipos. En rojo, se resalta el SNP también incluido en otros estudios (ver texto). Abreviaturas: i=intrón; e=exón.

En la fase de validación de nuestro estudio, estudiamos el tagSNP más significativo de la primera fase (rs2284734; OR=2.64; IC 95%=1.69-4.13; $P=1.8 \times 10^{-5}$) que, si bien no alcanzó la significación estadística en las series 2+3 (OR=1.42; IC 95%=0.99-2.03; $P=0.058$), mostró una tendencia en la misma dirección que en la primera fase, y bajo el mismo modelo de herencia recesivo. Esto sugiere que este *locus* debería ser analizado en detalle, bien en esta misma serie, o en un grupo independiente, buscando una variante funcional que explique la probable asociación de *TSHR* con CPT clásico. De hecho, el papel de este gen en CPT ha sido estudiado previamente, siendo frecuentes los cambios en la expresión y la presencia de mutaciones somáticas de *TSHR* en carcinomas diferenciados de tiroides (Garcia-Jimenez and Santisteban, 2007).

Por otra parte, varios estudios han descrito SNPs en *TSHR* asociados con una mayor susceptibilidad al desarrollo de patologías autoinmunes del tiroides, sobre todo enfermedad de Graves (Brand, et al., 2009; Gu, et al., 2010; Hiratani, et al., 2005), aceptándose así el papel de este *locus* como GBP en dicha enfermedad (Davies, et al., 2010). En uno de estos trabajos, se encontró una asociación muy significativa con enfermedad de Graves para el mismo tagSNP (rs2284734) elegido para la validación en nuestro estudio (Dechairo, et al., 2005), reforzando la importancia de este *locus*, al menos, en los procesos de inflamación del tiroides. Por el contrario, los estudios de asociación de *TSHR* en cáncer de tiroides han sido escasos, y hasta ahora, con resultados negativos (Lonn, et al., 2007; Matakidou, et al., 2004). Estos estudios se limitaron al estudio de variantes exónicas de *TSHR* en un grupo de 167 CPT asociados a radiación ocupacional (Lonn, et al., 2007), y en una serie más amplia, pero peor caracterizada, de 304 carcinomas diferenciados de tiroides (Matakidou, et al., 2004). Ambos estudios, al igual que el nuestro, no encontraron asociaciones significativas para las variantes rs1991517 (p.Asp727Glu) ni rs2075179 (p.Asn187Asn). Dado que sí obtuvimos una asociación muy significativa para otras variantes intrónicas de *TSHR* (figura D3), y teniendo en cuenta que la correlación entre variantes no es completa, es prematuro descartar la implicación de *TSHR*, y es probable que haga falta un mapeo más detallado de esta región, en busca de la variante funcional que explique la susceptibilidad a CPT clásico, y que muy posiblemente sea una variante intrónica y/o reguladora.

3.2. *TG*: gen de la tiroglobulina

Los resultados obtenidos para variantes en el gen de la tiroglobulina muestran algo similar a lo mencionado en *TSHR*, aunque quizá más complejo y menos consistente. La tiroglobulina, que actúa como precursor de las hormonas tiroideas, está codificada por un gen muy grande (48 exones, 268 kb) y con un patrón de LD complejo, en el que se obtuvieron resultados muy significativos en la serie 1 para varios polimorfismos (7 SNPs, $P < 0.01$), sugiriendo su implicación en la susceptibilidad a desarrollar CPTvf. Precisamente el fuerte LD en *TG* no permitió acotar una región concreta donde buscar una posible variante causal para validar de forma dirigida, de modo que se genotipó el mejor tagSNP de la serie 1 en las series 2+3, no detectándose ni siquiera una tendencia a la validación de este *locus*.

Trabajos previos han asociado polimorfismos en *TG* con un aumento en la susceptibilidad a desarrollar la enfermedad de Graves (Ban, et al., 2003; Ban, et al., 2004; Caputo, et al., 2010). De hecho, merece mención un estudio que asocia el polimorfismo rs1133076 (p.Arg2530Gln) en *TG* con el desarrollo de carcinomas de tiroides de estirpe folicular (OR=1.60; $P=0.015$) (Matakidou, et al., 2004). Este SNP, presente en nuestro estudio, no se asoció al desarrollo de CPTvf en la primera fase, pero sí lo hicieron otros en el mismo bloque de desequilibrio, por lo que quizá exista otra variante intrónica que pueda explicar la susceptibilidad a CPTvf. Por último, una publicación describió un ligamiento de la región 8q24 en familias con CPT y melanoma (He, et al., 2009). Tras acotar la región al máximo, los autores de este trabajo secuenciaron la región codificante íntegra de *TG*, incluido en dicha región, sin encontrar ninguna variante no descrita. Esto refuerza la idea de que, si este gen está implicado en la susceptibilidad a algún tipo de cáncer de tiroides, no parece que se deba a variantes exónicas.

3.3. El gen *SMAD7*

Los resultados obtenidos para el gen *SMAD7* son controvertidos. Este *locus* era *a priori* un buen candidato a jugar un papel en la susceptibilidad al desarrollo de CPT. Esto se debe a que Smad7 es un inhibidor esencial en la vía de señalización de TGF- β que, a su vez, juega un papel importante en el crecimiento y función del tiroides, y en el desarrollo de carcinomas de estirpe folicular (Pisarev, et al., 2009). En la serie 1, obtuvimos resultados significativos en CPT para 3 SNPs de *SMAD7*, además de para la distribución de haplotipos de dos bloques de LD en este *locus*. Para la validación, elegimos el tagSNP intrónico rs4939827 que parecía un buen marcador, ya que había sido asociado previamente en varios GWAS con cáncer colorrectal (Broderick, et al., 2007; Tenesa, et al., 2008; Tomlinson, et al., 2008). En estos estudios, el alelo rs4939827T confería un riesgo significativo al desarrollo de cáncer de colon. En cambio, en nuestras tres series de casos y controles, al igual que en la población HapMap-CEU, el alelo descrito como mayoritario es precisamente el alelo rs4939827T, considerándose el alelo C como polimórfico. Por este motivo, expresamos nuestros resultados en función del alelo menor C. Fueron llamativos los resultados significativos pero en direcciones opuestas que obtuvimos en nuestro estudio. Así, en la primera fase (serie 1), los portadores del genotipo rs4939827CC mostraron protección significativa al desarrollo de CPT (OR=0.68; IC 95%=0.50-0.94; $P=0.019$), mientras que en la fase de validación (series 2+3), ese mismo genotipo CC, bajo el mismo modelo recesivo, confería riesgo (OR=1.32; IC 95%=1.03-1.68; $P=0.027$). Un análisis más detallado de las series 2 y 3 por separado demostró que dicha variante no estaba asociada en la serie 2 (OR=1.09; IC 95%=0.75-1.58; $P=0.650$), pero confería riesgo en la serie 3 (OR=1.53; IC 95%=1.10-2.12; $P=0.011$). Esto sugiere una heterogeneidad entre ambas series para este *locus*, cosa que no sucedía para el resto de variantes. Lo realmente sorprendente es la tendencia opuesta entre las series 1 y 3, lo cual tiene una difícil interpretación biológica, y no puede explicarse como un efecto de subestructura poblacional entre series de muestras, ya que los individuos incluidos en la serie 3 tienen un origen similar al de la serie 1 (población española). Una posible solución, pendiente de confirmar en CPT, sería el genotipado de una nueva variante intrónica de *SMAD7*, que recientemente ha sido propuesta como la variante causal que explicaría el riesgo en cáncer colorrectal, mediante la alteración de la expresión de *SMAD7* (Pittman, et al., 2009).

3.4. Otros genes candidatos

Del resto de genes, cuyas variantes eran significativas en la serie 1, pero no se replicaron en la fase de validación, destaca el cambio de aminoácido rs1053266 en el gen *CCDC6*, que al menos mostró una tendencia en las series 2+3 (P no ajustada=0.070). Este gen, se transloca somáticamente con *RET* (RET/PTC1) en un porcentaje variable de CPT, siendo una de las alteraciones iniciadoras de estos tumores (Kondo, et al., 2006; Sugg, et al., 1998). Quedaría por elucidar el papel potencial que éste u otros SNPs de *CCDC6* tendrían en la susceptibilidad al desarrollo de CPT clásico. Una hipótesis, para la que serían necesarios experimentos específicos, sería que la variación germinal en *CCDC6* facilitase la capacidad de este gen para translocarse con *RET*, o que mimetizase en cierto grado la activación constitutiva que se produce en la translocación RET/PTC1. Para ello, haría falta un estudio con muestras pareadas (tejido tumoral y sangre periférica), no disponibles en nuestro estudio. Otra hipótesis sería que variantes de *CCDC6* interactuasen con el SNP rs2910164 del pre-miR-146a, recientemente asociado al desarrollo de CPT, y capaz de inhibir la expresión de *CCDC6* (Jazdzewski, et al., 2008). Para comprobar esta última suposición haría falta un estudio de epistasia, similar al realizado en nuestro estudio, que incluyese variantes de ambos genes, así como estudios funcionales concretos.

Respecto otras variantes consideradas como candidatas a influir en el desarrollo de CPT, las pertenecientes a los genes *PRKAR1A* y *KIT* podrían ser descartadas como factores de riesgo, al estar incluidas en análisis con potencia suficiente. Por el contrario, para la variante no sinónima de *STAT2*, dada su poca frecuencia (MAF=4 %), harían falta series mucho más amplias para llegar a resultados concluyentes.

De los polimorfismos en genes asociados a subtipos menos frecuentes de estas neoplasias (*GHR*, *PPARG*, *CSF2*), podemos extraer la idea de que aún hacen falta muchos estudios para abordar con la suficiente potencia estadística la evidente heterogeneidad de los carcinomas de tiroides de estirpe folicular. Resulta particularmente interesante que, aunque no validadas, encontrásemos variantes del gen *PPARG* asociadas a carcinomas con patrón de crecimiento folicular (CFT y CPTvf). Esto concuerda con el papel específico que este gen juega, a través de la translocación PAX8-PPARG, en el desarrollo de CFT (Kroll, et al., 2000) y de CPTvf (Castro, et al., 2006), y con el hecho de que CFT y CPTvf se agrupan juntos por patrones de expresión (Montero-Conde, et al., 2008). Análogamente a lo comentado para *CCDC6*,

quedaría por determinar cómo podrían influir las variantes germinales en lo descrito en tejido somático.

Respecto a las comparaciones en función de las características de los tumores, podríamos descartar el papel del cambio p.Ile117Thr de *CSF2* en la agresividad de los tumores, ya que no se validó ni en las series 2+3 del estudio caso-control, ni en la comparación entre casos metastásicos y aquéllos que no lo eran. Sin embargo, el gen *KIT*, si bien debería ser validado, podría ser un marcador de agresividad en estos tumores, lo cual encajaría con su papel descrito en la transformación maligna de tumores tiroideos (He, et al., 2005; Natali, et al., 1995).

3.5. El análisis de epistasia revela nuevos genes de susceptibilidad

En los últimos años, ante la gran cantidad de variantes asociadas a enfermedades complejas, se ha puesto en evidencia la necesidad de realizar análisis específicos para evaluar los efectos conjuntos. La importancia de estos análisis radica principalmente, en desvelar cómo la presencia de una variante genética puede condicionar el efecto de otra (Cordell, 2009; Dermitzakis and Clark, 2009).

Los análisis de epistasia, distintos de los estudios clásicos de efectos acumulados o de la identificación de modificadores del fenotipo, suponen un reto computacional y estadístico. Así, aunque hay varios métodos capaces de detectar interacciones entre genes, su rendimiento es desigual. En este sentido, el algoritmo MB-MDR (Calle, et al., 2008), empleado en nuestro trabajo, es un método muy potente, pero que computacionalmente sería inviable para una serie de datos procedente de un GWAS, en cambio siendo aplicable a un número de comparaciones como el de nuestro estudio. Además, fue seleccionado por su capacidad para corregir los resultados en varios niveles (sustracción de los efectos individuales, inclusión de cofactores, tests de permutaciones,...). El hecho de haber sido capaces de validar algunas de las interacciones candidatas tras los ajustes pertinentes, refuerza la elección de este método y la propia asociación de estas parejas de genes.

Hasta la fecha, los estudios en levadura han sido útiles para la detección y puesta a punto de análisis epistáticos (Boone, et al., 2007; Brem, et al., 2005; Snitkin and Segre, 2011). Por el contrario, los intentos de detección de interacciones relacionadas con la susceptibilidad a desarrollar enfermedades en humanos han sido

escasos, y sólo muy recientemente han empezado a dar resultados (Kong, et al., 2011; Naushad, et al., 2010; Wu, et al., 2010), aunque necesitarían validaciones independientes. En este sentido, el hecho de haber detectado y validado en nuestro trabajo efectos epistásicos entre varias parejas de genes relacionados con la susceptibilidad a desarrollar CPT es notable.

Los resultados del análisis de epistasia en nuestro grupo de CPT frente a controles revelaron un papel importante del gen *PAX8*. Este factor de transcripción es conocido por participar en el desarrollo del tiroides, relacionarse con enfermedades congénitas de esta glándula y regular la expresión de genes específicos de la célula folicular tiroidea (Macchia, et al., 1998; Mansouri, et al., 1998; Ohno, et al., 1999; Zannini, et al., 1992). Según nuestros resultados, *PAX8* podría interactuar con los genes *KIT*, *STAT2* y *STK17B*. Aunque no está descrito que *PAX8* regule la expresión de estos genes, sería interesante evaluarlo con ensayos específicamente destinados a ello. De hecho, *STK17B*, por ejemplo, se eligió por estar diferencialmente expresado al comparar CPT frente a tejido normal, por lo que una diferencia transcripcional determinada por la acción de *PAX8* podría ser el mecanismo subyacente. Sin embargo, queda por determinar hasta qué punto las interacciones entre genes identificadas en nuestro trabajo pueden ser útiles como marcadores de riesgo en CPT.

4. Factores genéticos de riesgo al cáncer de tiroides. ¿Hemos avanzado en su conocimiento?

La susceptibilidad genética al desarrollo de carcinomas de tiroides de células foliculares ha experimentado sin duda un avance importante en los últimos años, aunque aún falta mucho para explicar la forma en la que el fondo genético de cada individuo determina su riesgo a desarrollar la enfermedad.

Además de asociaciones ya mencionadas al principio de esta sección, hay un pequeño número de asociaciones, identificadas en estudios con potencia suficiente y/o validados en series independientes. En este sentido, podríamos considerar que los factores genéticos más consistentemente identificados son el rs2910164, localizado en el pre-miR-146a (Jazdzewski, et al., 2008), el rs25487 (p.Arg399Gln) en *XRCC1* (Akulevich, et al., 2009; Ho, et al., 2009), el rs1133076 (p.Arg2530Gln) en *TG* (Matakidou, et al., 2004), el rs4752904 (p.Asp872Glu) en *PTPRJ* (Iuliano, et al., 2010), el rs861539 (p.Thr241Met) en *XRCC3* (Bastos, et al., 2009; Sturgis, et al., 2005), el rs1042522 (p.Arg72Pro) en *TP53* (Akulevich, et al., 2009; Granja, et al., 2004; Rogounovitch, et al., 2006), y las variantes intergénicas rs944289 (Gudmundsson, et al., 2009) y rs6983267 (Wokolorczyk, et al., 2008) en 14q13.3 y 8q24.21, respectivamente. Como se muestra en la figura D4, nuestro trabajo ha contribuido decisivamente a desvelar nuevos factores de riesgo en el desarrollo de carcinomas de tiroides de estirpe folicular, con la identificación de las variantes rs34330 en *CDKN1B* (Landa, et al., 2010) y rs1867277 en *FOXE1* (Landa, et al., 2009).

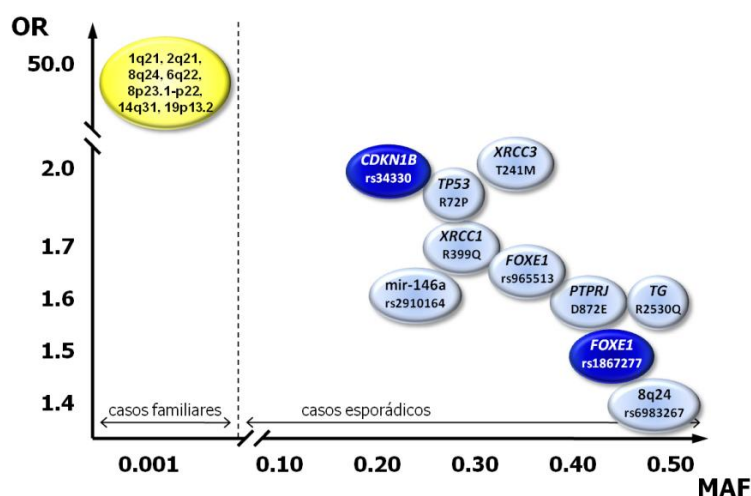


Figura D4. Representación esquemática de la contribución de variantes genéticas al desarrollo de carcinomas de tiroides de estirpe folicular, según el riesgo (OR) atribuible a cada variante y su frecuencias alélica (MAF). Sólo se muestran los GBP más consistentemente

asociados al desarrollo de cáncer de tiroides y, entre ellos, en azul oscuro, se resaltan los factores genéticos de riesgo identificados en esta tesis. Abreviaturas: OR= *odds ratio*; MAF= *Minor Allele Frequency*.

Estos estudios han señalado genes implicados fundamentalmente en el metabolismo de la célula folicular tiroidea, en rutas de reparación del ADN y en el control del ciclo celular. La implicación de estas rutas en concreto refuerza la idea de que cuando se dispone de un conocimiento biológico previo, los estudios de asociación dirigidos a genes o vías metabólicas específicas pueden ser útiles en la identificación de GBP. Por el contrario, los GWAS, si bien tienen la ventaja de que, al ser una aproximación libre de hipótesis, permiten identificar genes antes desconocidos, tiene la limitación de que aproximadamente el 20 % de los SNPs están capturados parcialmente, mientras que las variantes raras no lo están en absoluto.

A la vista de nuestros resultados en la primera fase del estudio (figura R6), aunque se observa una tendencia lógica hacia una mayor tasa de asociaciones significativas conforme aumenta la frecuencia de la variante estudiada ($MAF > 25\%$), parece ser que nuestro diseño, al estar dirigido hacia un número más bien reducido de genes y SNPs interesantes *a priori*, proporciona una tasa de éxito bastante similar a lo largo de todo el rango de MAFs. Sin embargo, reconocemos que existe la posibilidad de que esta observación esté condicionada por los genes elegidos, el tipo de variantes elegidas en cada gen, y la frecuencia de éstas. Además, como puede observarse en la figura D4, los genes de susceptibilidad mejor caracterizados hasta ahora tienen un MAF siempre mayor del 20 %, lo que hace pensar que nadie ha sido capaz aún de diseñar un estudio con la potencia suficiente para desvelar la contribución genética de variantes más raras, especialmente aquéllas con $MAF < 10\%$. Esto podría explicar, en parte, la gran proporción del riesgo atribuible que aún no ha sido explicada. Una nueva corriente de pensamiento acude a los métodos de resecuenciación para identificar variantes raras que difícilmente estarán capturadas por los SNPs incluidos en los GWAS. Sin embargo, es fácil calcular que el tamaño muestral necesario para demostrar asociación con variantes muy raras probablemente supere la realidad de cualquier estudio. Otra posibilidad es explorar la interacción gen-gen, que es por el momento un capítulo pendiente, a la espera de que existan programas capaces de manejar la información en un tiempo razonable, y con un uso razonable de recursos. Este tipo de estudios de epistasia posiblemente sólo puedan aplicarse a un conjunto de genes limitado, como por ejemplo, todos aquellos que pertenezcan a una ruta metabólica, y no a todos los genes contenidos en un GWAS. Considerando solo genes relacionados de esta forma, añadimos un sentido biológico para la interacción, y nos devuelve al uso de una aproximación de genes candidatos.

La falta de series de pacientes suficientemente amplias es un problema común en los estudios de asociación de muchas enfermedades complejas (Houlston and Peto, 2004; Kere, 2010; McCarthy, et al., 2008), en especial en aquellas relativamente poco frecuentes, como son los carcinomas de tiroides. Nuestro estudio supone un esfuerzo para superar esta limitación, ya que ha conseguido reclutar la serie más amplia de esta enfermedad estudiada hasta la fecha (n=1629 casos). Sin embargo, es evidente que el establecimiento de consorcios internacionales sería de gran utilidad. Esto ayudaría, mediante una reevaluación de lo publicado en este campo, a establecer qué asociaciones son reales y qué riesgo confieren al desarrollo de carcinomas de tiroides.

Otra cuestión importante en este campo, que hemos tratado de resolver en nuestro estudio, es la heterogeneidad de esta neoplasia, que hasta ahora apenas ha sido considerada. La mayor parte de las publicaciones describen asociaciones en “carcinomas de tiroides no medulares”, o en CPT, por ser el subtipo predominante. Nuestros hallazgos demuestran que la estratificación de los pacientes en subgrupos de características homogéneas y bien establecidas es crucial. Parece lógico que, si los CPT, los CFT, y cada subtipo de CPT (clásico, variante folicular, etc.) son clínicamente distintos y, en algunos casos, muestran alteraciones somáticas diferentes, también impliquen alteraciones germinales o GBP distintivos. Es obvio también que el tratamiento de la heterogeneidad saldría beneficiado del establecimiento de grupos internacionales de colaboración.

Por otra parte, es conocido que en las enfermedades complejas o multifactoriales, el reto principal es elucidar el vínculo entre las variantes asociadas y el fenotipo observado (Frazer, et al., 2009; Katsanis, 2009). En este sentido, hemos intentado, en la medida de lo posible, realizar ensayos funcionales para establecer el papel funcional y el posible mecanismo por el que serían importantes determinados genes. Sin embargo, harían falta aún más ensayos para mejorar el conocimiento funcional de los genes identificados por nosotros y por otros grupos. El camino entre una señal de asociación y la variante funcional casi nunca es sencillo, pero resulta esencial para entender la biología del cáncer, y potencialmente mejorar su pronóstico y tratamiento. En este sentido, sería interesante aprovechar las nuevas tecnologías de secuenciación masiva, y resecuenciar aquellas regiones o genes asociados con carcinomas de tiroides, pero todavía no explicados desde un punto de vista funcional.

Por último, y teniendo en cuenta la naturaleza multifactorial de estas neoplasias, parece obvio que no pueden ser explicadas por unas pocas variantes, bastante frecuentes en población general, y genes aislados. Por eso, es necesario establecer cómo interactúan entre sí los genes identificados. Una forma de expresar esta interacción es la contribución epistática, es decir, el efecto fenotípico condicional que varios polimorfismos generan, que no es resultado de la suma de sus efectos individuales (Dimas and Dermitzakis, 2009). En nuestro trabajo hemos intentado descifrar estas contribuciones, y hemos podido identificar parejas de genes que condicionarían la susceptibilidad a desarrollar carcinomas de tiroides. Sin embargo, estos estudios deberían ser ampliados en los próximos años, incluyendo otros genes y relacionándolos con factores ambientales, para llegar a una visión más integrativa de esta enfermedad, al modo de lo representado en la figura D5.

En resumen, nuestro trabajo prueba que, cuando se dispone de conocimiento sobre la biología de la enfermedad, además de una información clínica de confianza, los estudios de asociación que adoptan una aproximación de genes candidatos pueden ser métodos directos y sencillos para identificar genes de baja penetrancia y para encontrar las posibles variantes funcionales en esos genes.

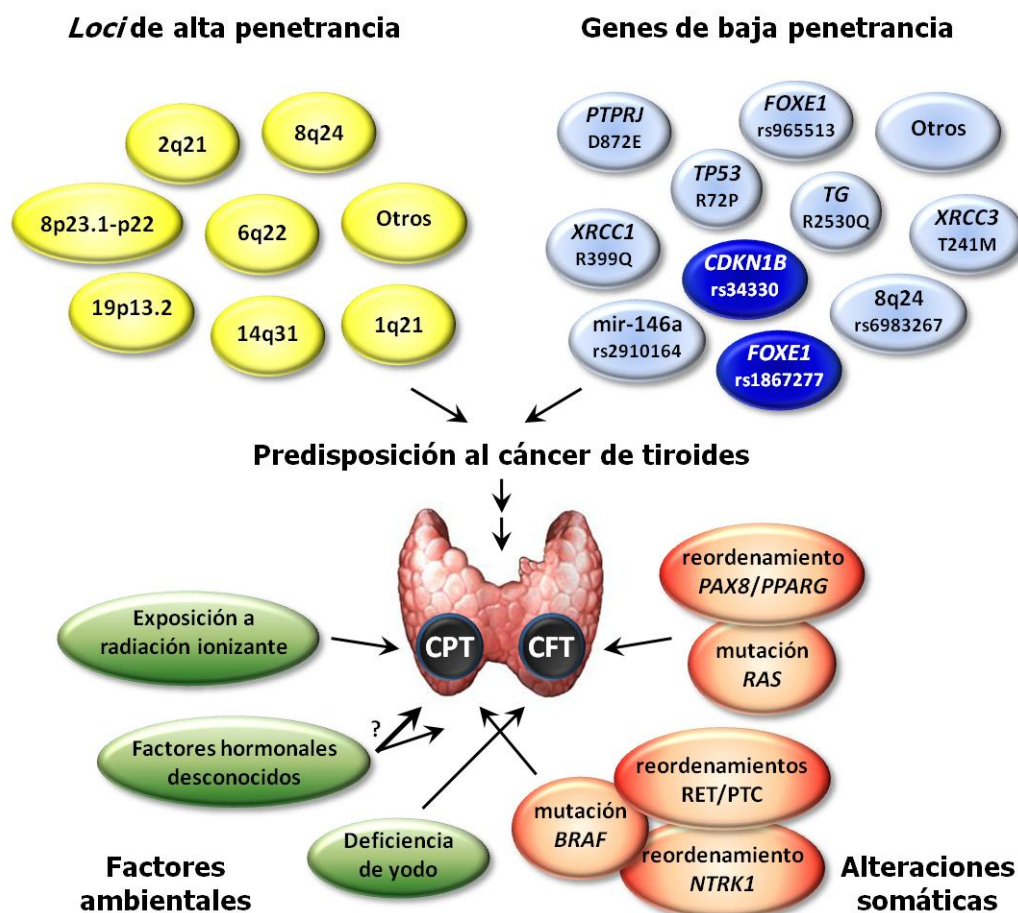
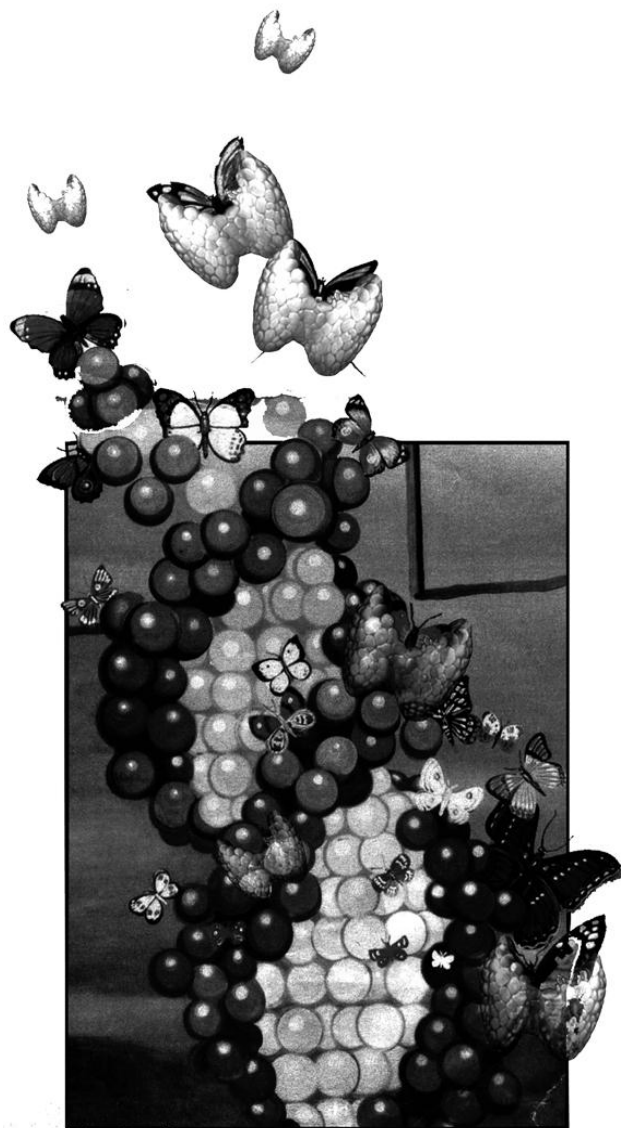


Figura D5. Modelo de heterogeneidad genética y ambiental en la predisposición individual al desarrollo de carcinomas de tiroides de estirpe folicular. En este modelo, son necesarios varios factores para el desarrollo de CPT o CFT. El fondo genético de cada individuo, bien monogénico en los casos familiares, o poligénico en los esporádicos, determina una condición de predisposición al cáncer de tiroides para un individuo concreto. Además, los agentes exógenos dañinos, y la adquisición de mutaciones somáticas deletéreas en genes específicos, determinan en última instancia el desarrollo de la enfermedad. Sólo se muestran los GBP más consistentemente asociados al desarrollo de cáncer de tiroides y, entre ellos, en azul oscuro, se resaltan los factores genéticos de riesgo identificados en esta tesis.

CONCLUSIONES



I. Los estudios de asociación basados en la selección de genes candidatos, cuando están fundamentados en un conocimiento biológico sólido, son una herramienta válida para la detección de genes de baja penetrancia (GBP) en carcinomas de tiroides de estirpe folicular.

II. La estratificación de los pacientes en subgrupos de características homogéneas es un paso necesario y esencial para abordar la heterogeneidad de esta enfermedad y detectar GBP específicamente asociados a cada subtipo histológico.

III. El polimorfismo rs34330 (c.-79C>T) del gen *CDKN1B* (p27) confiere riesgo a desarrollar la variante folicular del carcinoma papilar de tiroides (CPTvf) en población española (OR=2.12; IC 95%=1.09-4.15; $P=0.023$). Los portadores del genotipo de riesgo rs34330TT muestran una tasa de transcripción menor de *CDKN1B*, lo que supone un mecanismo alternativo de regulación negativa de p27.

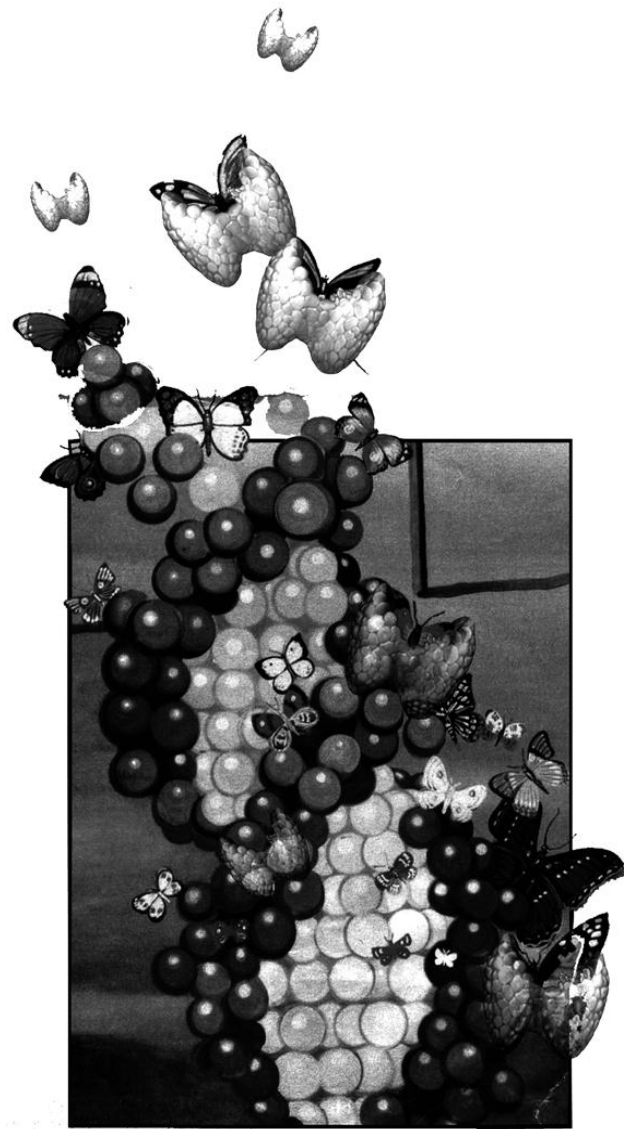
IV. La variante rs1867277 (c.-283G>A) del gen *FOXE1* está asociada de forma altamente significativa con el desarrollo del carcinoma papilar de tiroides (CPT) en tres poblaciones caucásicas independientes (OR (por alelo)=1.45; IC 95 % =1.30-1.61; $P=4.7 \times 10^{-12}$). El mecanismo subyacente que explicaría la susceptibilidad es el reclutamiento diferencial de los factores de transcripción USF1 y USF2, que se produce exclusivamente en presencia del alelo de riesgo rs1867277A.

V. Los genes *TSHR*, *TG* y *SMAD7* son candidatos a actuar como GBP en carcinomas de células foliculares de tiroides, si bien todavía no se han identificado las posibles variantes causales en cada uno de ellos.

VI. El estudio de interacción gen-gen o epistasia, además de proporcionar una visión más completa de la susceptibilidad genética al cáncer de tiroides, ha revelado nuevos GBP potenciales, siendo de especial interés la implicación del gen *PAX8*.

VII. Gran parte de la contribución genética al desarrollo de carcinomas de tiroides de célula folicular está aún por identificar. Para avanzar en su conocimiento, en el futuro serán necesarios: 1. el establecimiento de consorcios internacionales, capaces de reclutar un gran número de pacientes; 2. la obtención de datos clínicos precisos que permitan estratificar la muestra; 3. la identificación y caracterización funcional de las variantes causales en cada gen; y 4. el desarrollo de herramientas que permitan detectar con fiabilidad interacciones entre varios genes, así como identificar variantes menos frecuentes en la población.

BIBLIOGRAFÍA



- Adjadj E, Schlumberger M, de Vathaire F. 2009. Germ-line DNA polymorphisms and susceptibility to differentiated thyroid cancer. *Lancet Oncol* 10(2):181-90.
- Akulevich NM, Saenko VA, Rogounovitch TI, Drozd VM, Lushnikov EF, Ivanov VK, Mitsutake N, Kominami R, Yamashita S. 2009. Polymorphisms of DNA damage response genes in radiation-related and sporadic papillary thyroid carcinoma. *Endocr Relat Cancer* 16(2):491-503.
- Andrews NC, Faller DV. 1991. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res* 19(9):2499.
- Aza-Blanc P, Di Lauro R, Santisteban P. 1993. Identification of a cis-regulatory element and a thyroid-specific nuclear factor mediating the hormonal regulation of rat thyroid peroxidase promoter activity. *Mol Endocrinol* 7(10):1297-306.
- Ban Y, Greenberg DA, Concepcion E, Skrabanek L, Villanueva R, Tomer Y. 2003. Amino acid substitutions in the thyroglobulin gene are associated with susceptibility to human and murine autoimmune thyroid disease. *Proc Natl Acad Sci U S A* 100(25):15119-24.
- Ban Y, Tozaki T, Taniyama M, Tomita M. 2004. Association of a thyroglobulin gene polymorphism with Hashimoto's thyroiditis in the Japanese population. *Clin Endocrinol (Oxf)* 61(2):263-8.
- Barrans SL, Fenton JA, Banham A, Owen RG, Jack AS. 2004. Strong expression of FOXP1 identifies a distinct subset of diffuse large B-cell lymphoma (DLBCL) patients with poor outcome. *Blood* 104(9):2933-5.
- Bastos HN, Antao MR, Silva SN, Azevedo AP, Manita I, Teixeira V, Pina JE, Gil OM, Ferreira TC, Limbert E and others. 2009. Association of polymorphisms in genes of the homologous recombination DNA repair pathway and thyroid cancer risk. *Thyroid* 19(10):1067-75.
- Beimfohr C, Klugbauer S, Demidchik EP, Lengfelder E, Rabes HM. 1999. NTRK1 re-arrangement in papillary thyroid carcinomas of children after the Chernobyl reactor accident. *Int J Cancer* 80(6):842-7.
- Bignell GR, Canzian F, Shayeghi M, Stark M, Shugart YY, Biggs P, Mangion J, Hamoudi R, Rosenblatt J, Buu P and others. 1997. Familial nontoxic multinodular thyroid goiter locus maps to chromosome 14q but does not account for familial nonmedullary thyroid cancer. *Am J Hum Genet* 61(5):1123-30.
- Bongarzone I, Fugazzola L, Vigneri P, Mariani L, Mondellini P, Pacini F, Basolo F, Pinchera A, Pilotti S, Pierotti MA. 1996. Age-related activation of the tyrosine kinase receptor protooncogenes RET and NTRK1 in papillary thyroid carcinoma. *J Clin Endocrinol Metab* 81(5):2006-9.
- Boone C, Bussey H, Andrews BJ. 2007. Exploring genetic interactions and networks with yeast. *Nat Rev Genet* 8(6):437-49.
- Bounacer A, Wicker R, Caillou B, Cailleux AF, Sarasin A, Schlumberger M, Suarez HG. 1997. High prevalence of activating ret proto-oncogene rearrangements, in thyroid tumors from patients who had received external radiation. *Oncogene* 15(11):1263-73.
- Brand OJ, Barrett JC, Simmonds MJ, Newby PR, McCabe CJ, Bruce CK, Kysela B, Carr-Smith JD, Brix T, Hunt PJ and others. 2009. Association of the thyroid stimulating hormone receptor gene (TSHR) with Graves' disease. *Hum Mol Genet* 18(9):1704-13.
- Brem RB, Storey JD, Whittle J, Kruglyak L. 2005. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* 436(7051):701-3.
- Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S and others. 2007. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet* 39(11):1315-7.
- Bufalo NE, Leite JL, Guilhen AC, Morari EC, Granja F, Assumpcao LV, Ward LS. 2006. Smoking and susceptibility to thyroid cancer: an inverse association with CYP1A1 allelic variants. *Endocr Relat Cancer* 13(4):1185-93.
- Calle ML, Urrea V, Vellalta G, Malats N, Steen KV. 2008. Improving strategies for detecting genetic patterns of disease susceptibility in association studies. *Stat Med* 27(30):6532-46.
- Canzian F, Amati P, Harach HR, Kraimps JL, Lesueur F, Barbier J, Levillain P, Romeo G, Bonneau D. 1998. A gene predisposing to familial thyroid tumors with cell oxyphilia maps to chromosome 19p13.2. *Am J Hum Genet* 63(6):1743-8.
- Capezzone M, Marchisotta S, Cantara S, Busonero G, Brilli L, Pazaitou-Panayiotou K, Carli AF, Caruso G, Toti P, Capitani S and others. 2008. Familial non-medullary thyroid carcinoma

- displays the features of clinical anticipation suggestive of a distinct biological entity. *Endocr Relat Cancer* 15(4):1075-81.
- Caputo M, Rivolta CM, Mories T, Corrales JJ, Galindo P, Gonzalez-Sarmiento R, Targovnik HM, Miralles-Garcia JM. 2010. Analysis of thyroglobulin gene polymorphisms in patients with autoimmune thyroiditis. *Endocrine* 37(3):389-95.
- Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N and others. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22(3):231-8.
- Carrion AM, Link WA, Ledo F, Mellstrom B, Naranjo JR. 1999. DREAM is a Ca²⁺-regulated transcriptional repressor. *Nature* 398(6722):80-4.
- Castro P, Rebocho AP, Soares RJ, Magalhaes J, Roque L, Trovisco V, Vieira de Castro I, Cardoso-de-Oliveira M, Fonseca E, Soares P and others. 2006. PAX8-PPARgamma rearrangement is frequently detected in the follicular variant of papillary thyroid carcinoma. *J Clin Endocrinol Metab* 91(1):213-20.
- Cavaco BM, Batista PF, Sobrinho LG, Leite V. 2008. Mapping a new familial thyroid epithelial neoplasia susceptibility locus to chromosome 8p23.1-p22 by high-density single-nucleotide polymorphism genome-wide linkage analysis. *J Clin Endocrinol Metab* 93(11):4426-30.
- Conde L, Vaquerizas JM, Dopazo H, Arbiza L, Reumers J, Rousseau F, Schymkowitz J, Dopazo J. 2006. PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res* 34(Web Server issue):W621-5.
- Consortium IH. 2003. The International HapMap Project. *Nature* 426(6968):789-96.
- Cordell HJ. 2009. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet* 10(6):392-404.
- Corre S, Galibert MD. 2005. Upstream stimulating factors: highly versatile stress-responsive transcription factors. *Pigment Cell Res* 18(5):337-48.
- Coulson JM, Edgson JL, Marshall-Jones ZV, Mulgrew R, Quinn JP, Woll PJ. 2003. Upstream stimulatory factor activates the vasopressin promoter via multiple motifs, including a non-canonical E-box. *Biochem J* 369(Pt 3):549-61.
- Cuesta I, Zaret KS, Santisteban P. 2007. The forkhead factor FoxE1 binds to the thyroperoxidase promoter during thyroid cell differentiation and modifies compacted chromatin structure. *Mol Cell Biol* 27(20):7302-14.
- Curiel TJ, Coukos G, Zou L, Alvarez X, Cheng P, Mottram P, Evdemon-Hogan M, Conejo-Garcia JR, Zhang L, Burow M and others. 2004. Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. *Nat Med* 10(9):942-9.
- Cybulski C, Gorski B, Huzarski T, Masojc B, Mierzejewski M, Debniak T, Teodorczyk U, Byrski T, Gronwald J, Matyjasik J and others. 2004. CHEK2 is a multiorgan cancer susceptibility gene. *Am J Hum Genet* 75(6):1131-5.
- Czene K, Lichtenstein P, Hemminki K. 2002. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer* 99(2):260-6.
- Chakravarti A. 1999. Population genetics--making sense out of sequence. *Nat Genet* 21(1 Suppl):56-60.
- Chang BL, Zheng SL, Isaacs SD, Wiley KE, Turner A, Li G, Walsh PC, Meyers DA, Isaacs WB, Xu J. 2004. A polymorphism in the CDKN1B gene is associated with increased risk of hereditary prostate cancer. *Cancer Res* 64(6):1997-9.
- Chen AY, Jemal A, Ward EM. 2009. Increasing incidence of differentiated thyroid cancer in the United States, 1988-2005. *Cancer* 115(16):3801-7.
- Chiappetta G, De Marco C, Quintiero A, Califano D, Gherardi S, Malanga D, Scrima M, Montero-Conde C, Cito L, Monaco M and others. 2007. Overexpression of the S-phase kinase-associated protein 2 in thyroid cancer. *Endocr Relat Cancer* 14(2):405-20.
- Chu IM, Hengst L, Slingerland JM. 2008. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nat Rev Cancer* 8(4):253-67.
- D'Andrea B, Di Palma T, Mascia A, Motti ML, Viglietto G, Nitsch L, Zannini M. 2005. The transcriptional repressor DREAM is involved in thyroid gene expression. *Exp Cell Res* 305(1):166-78.

- Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W and others. 2002. Mutations of the BRAF gene in human cancer. *Nature* 417(6892):949-54.
- Davies L, Welch HG. 2006. Increasing incidence of thyroid cancer in the United States, 1973-2002. *JAMA* 295(18):2164-7.
- Davies TF, Yin X, Latif R. 2010. The genetics of the thyroid stimulating hormone receptor: history and relevance. *Thyroid* 20(7):727-36.
- De Felice M, Di Lauro R. 2004. Thyroid development and its disorders: genetics and molecular mechanisms. *Endocr Rev* 25(5):722-46.
- De Felice M, Ovitt C, Biffali E, Rodriguez-Mallon A, Arra C, Anastassiadis K, Macchia PE, Mattei MG, Mariano A, Scholer H and others. 1998. A mouse model for hereditary thyroid dysgenesis and cleft palate. *Nat Genet* 19(4):395-8.
- Dechairo BM, Zabaneh D, Collins J, Brand O, Dawson GJ, Green AP, Mackay I, Franklyn JA, Connell JM, Wass JA and others. 2005. Association of the TSHR gene with Graves' disease: the first disease specific locus. *Eur J Hum Genet* 13(11):1223-30.
- DeLellis RA. 2004. Pathology and genetics of tumours of endocrine organs. Lyon: IARC Press.
- Dermitzakis ET, Clark AG. 2009. Genetics. Life after GWA studies. *Science* 326(5950):239-40.
- Dimas AS, Dermitzakis ET. 2009. Genetic variation of regulatory systems. *Curr Opin Genet Dev* 19(6):586-90.
- Domarus Av, Farreras Valentí P, Rozman C. 1988. Medicina interna. Barcelona: Doyma.
- Dong LM, Potter JD, White E, Ulrich CM, Cardon LR, Peters U. 2008. Genetic susceptibility to cancer: the role of polymorphisms in candidate genes. *JAMA* 299(20):2423-36.
- Driver KE, Song H, Lesueur F, Ahmed S, Barbosa-Morais NL, Tyrer JP, Ponder BA, Easton DF, Pharoah PD, Dunning AM. 2008. Association of single-nucleotide polymorphisms in the cell cycle genes with breast cancer in the British population. *Carcinogenesis* 29(2):333-41.
- Dupont WD, Plummer WD, Jr. 1998. Power and sample size calculations for studies involving linear regression. *Control Clin Trials* 19(6):589-601.
- Eichberger T, Regl G, Ikram MS, Neill GW, Philpott MP, Aberger F, Frischauf AM. 2004. FOXE1, a new transcriptional target of GLI2 is expressed in human epidermis and basal cell carcinoma. *J Invest Dermatol* 122(5):1180-7.
- Erickson LA, Jin L, Wollan PC, Thompson GB, van Heerden J, Lloyd RV. 1998. Expression of p27kip1 and Ki-67 in benign and malignant thyroid tumors. *Mod Pathol* 11(2):169-74.
- Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*.
- Fletcher O, Houlston RS. 2010. Architecture of inherited susceptibility to common cancer. *Nat Rev Cancer* 10(5):353-61.
- Foulkes NS, Borrelli E, Sassone-Corsi P. 1991. CREM gene: use of alternative DNA-binding domains generates multiple antagonists of cAMP-induced transcription. *Cell* 64(4):739-49.
- Frazer KA, Murray SS, Schork NJ, Topol EJ. 2009. Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 10(4):241-51.
- Fujimoto Y, Obara T, Ito Y, Kodama T, Aiba M, Yamaguchi K. 1990. Diffuse sclerosing variant of papillary carcinoma of the thyroid. Clinical importance, surgical treatment, and follow-up study. *Cancer* 66(11):2306-12.
- Galvan A, Ioannidis JP, Dragani TA. 2010. Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. *Trends Genet* 26(3):132-41.
- Garcia-Jimenez C, Santisteban P. 2007. TSH signalling and cancer. *Arq Bras Endocrinol Metabol* 51(5):654-71.
- Gaspar J, Rodrigues S, Gil OM, Manita I, Ferreira TC, Limbert E, Goncalves L, Pina JE, Rueff J. 2004. Combined effects of glutathione S-transferase polymorphisms and thyroid cancer risk. *Cancer Genet Cytogenet* 151(1):60-7.
- Gayther SA, Song H, Ramus SJ, Kjaer SK, Whittemore AS, Quaye L, Tyrer J, Shadforth D, Hogdall E, Hogdall C and others. 2007. Tagging single nucleotide polymorphisms in cell cycle control genes and susceptibility to invasive epithelial ovarian cancer. *Cancer Res* 67(7):3027-35.
- Ghossein R, Livolsi VA. 2008. Papillary thyroid carcinoma tall cell variant. *Thyroid* 18(11):1179-81.

- Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. 1994. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst* 86(21):1600-8.
- Gonzalez P, Diez-Juan A, Coto E, Alvarez V, Reguero JR, Batalla A, Andres V. 2004. A single-nucleotide polymorphism in the human p27kip1 gene (-838C>A) affects basal promoter activity and the risk of myocardial infarction. *BMC Biol* 2:5.
- Granja F, Morari J, Morari EC, Correa LA, Assumpcao LV, Ward LS. 2004. Proline homozygosity in codon 72 of p53 is a factor of susceptibility for thyroid cancer. *Cancer Lett* 210(2):151-7.
- Greco A, Pierotti MA, Bongarzone I, Pagliardini S, Lanzi C, Della Porta G. 1992. TRK-T1 is a novel oncogene formed by the fusion of TPR and TRK genes in human papillary thyroid carcinomas. *Oncogene* 7(2):237-42.
- Grieco M, Cerrato A, Santoro M, Fusco A, Melillo RM, Vecchio G. 1994. Cloning and characterization of H4 (D10S170), a gene involved in RET rearrangements in vivo. *Oncogene* 9(9):2531-5.
- Gu LQ, Zhu W, Zhao SX, Zhao L, Zhang MJ, Cui B, Song HD, Ning G, Zhao YJ. 2010. Clinical associations of the genetic variants of CTLA-4, Tg, TSHR, PTPN22, PTPN12 and FCRL3 in patients with Graves' disease. *Clin Endocrinol (Oxf)* 72(2):248-55.
- Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, He H, Blondal T, Geller F, Jakobsdottir M and others. 2009. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat Genet*.
- Harach HR, Soubeyran I, Brown A, Bonneau D, Longy M. 1999. Thyroid pathologic findings in patients with Cowden disease. *Ann Diagn Pathol* 3(6):331-40.
- He H, Jazdzewski K, Li W, Liyanarachchi S, Nagy R, Volinia S, Calin GA, Liu CG, Franssila K, Suster S and others. 2005. The role of microRNA genes in papillary thyroid carcinoma. *Proc Natl Acad Sci U S A* 102(52):19075-80.
- He H, Nagy R, Liyanarachchi S, Jiao H, Li W, Suster S, Kere J, de la Chapelle A. 2009. A susceptibility locus for papillary thyroid carcinoma on chromosome 8q24. *Cancer Res* 69(2):625-31.
- Hemminki K, Rawal R, Chen B, Bermejo JL. 2004. Genetic epidemiology of cancer: from families to heritable genes. *Int J Cancer* 111(6):944-50.
- Hernandez A, Cespedes W, Xamena N, Surrallés J, Creus A, Galofre P, Marcos R. 2003. Glutathione S-transferase polymorphisms in thyroid cancer patients. *Cancer Lett* 190(1):37-44.
- Hernandez A, Xamena N, Surrallés J, Galofre P, Velazquez A, Creus A, Marcos R. 2008. Role of GST and NAT2 polymorphisms in thyroid cancer. *J Endocrinol Invest* 31(11):1025-31.
- Heselmeyer K, Macville M, Schrock E, Blegen H, Hellstrom AC, Shah K, Auer G, Ried T. 1997. Advanced-stage cervical carcinomas are defined by a recurrent pattern of chromosomal aberrations revealing high genetic instability and a consistent gain of chromosome arm 3q. *Genes Chromosomes Cancer* 19(4):233-40.
- Hilger-Eversheim K, Moser M, Schorle H, Buettner R. 2000. Regulatory roles of AP-2 transcription factors in vertebrate development, apoptosis and cell-cycle control. *Gene* 260(1-2):1-12.
- Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 106(23):9362-7.
- Hiratani H, Bowden DW, Ikegami S, Shirasawa S, Shimizu A, Iwatani Y, Akamizu T. 2005. Multiple SNPs in intron 7 of thyrotropin receptor are associated with Graves' disease. *J Clin Endocrinol Metab* 90(5):2898-903.
- Ho T, Li G, Lu J, Zhao C, Wei Q, Sturgis EM. 2009. Association of XRCC1 polymorphisms and risk of differentiated thyroid carcinoma: a case-control analysis. *Thyroid* 19(2):129-35.
- Ho T, Li G, Zhao C, Wei Q, Sturgis EM. 2005. RET polymorphisms and haplotypes and risk of differentiated thyroid cancer. *Laryngoscope* 115(6):1035-41.
- Hou P, Liu D, Shan Y, Hu S, Studeman K, Condouris S, Wang Y, Trink A, El-Naggar AK, Tallini G and others. 2007. Genetic alterations and their relationship in the phosphatidylinositol 3-kinase/Akt pathway in thyroid cancer. *Clin Cancer Res* 13(4):1161-70.
- Houlston RS, Peto J. 2004. The search for low-penetrance cancer susceptibility alleles. *Oncogene* 23(38):6471-6.

- Hsiao PJ, Lu MY, Chiang FY, Shin SJ, Tai YD, Juo SH. 2007. Vascular endothelial growth factor gene polymorphisms in thyroid cancer. *J Endocrinol* 195(2):265-70.
- Ioannidis JP, Castaldi P, Evangelou E. 2010. A compendium of genome-wide associations for cancer: critical synopsis and reappraisal. *J Natl Cancer Inst* 102(12):846-58.
- Ito E, Iwahashi Y, Yanagisawa Y, Suzuki Y, Sugano S, Yuasa Y, Maruyama K. 1999. Two short sequences have positive effects on the human p27Kip1 gene transcription. *Gene* 228(1-2):93-100.
- Iuliano R, Palmieri D, He H, Iervolino A, Borbone E, Pallante P, Cianflone A, Nagy R, Alder H, Calin GA and others. 2010. Role of PTPRJ genotype in papillary thyroid carcinoma risk. *Endocr Relat Cancer* 17(4):1001-6.
- Jacobson DA, Cho J, Landa LR, Jr., Tamarina NA, Roe MW, Buxbaum JD, Philipson LH. 2006. Downstream regulatory element antagonistic modulator regulates islet prodynorphin expression. *Am J Physiol Endocrinol Metab* 291(3):E587-95.
- Jazdzewski K, Murray EL, Franssila K, Jarzab B, Schoenberg DR, de la Chapelle A. 2008. Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proc Natl Acad Sci U S A* 105(20):7269-74.
- Joyce T, Cantarella D, Isella C, Medico E, Pintzas A. 2009. A molecular signature for Epithelial to Mesenchymal transition in a human colon cancer cell system is revealed by large-scale microarray analysis. *Clin Exp Metastasis*.
- Katsanis N. 2009. From association to causality: the new frontier for complex traits. *Genome Med* 1(2):23.
- Kere J. 2010. Genetics of complex disorders. *Biochem Biophys Res Commun* 396(1):143-6.
- Khan A, Smellie J, Nutting C, Harrington K, Newbold K. 2010. Familial nonmedullary thyroid cancer: a review of the genetics. *Thyroid* 20(7):795-801.
- Khoo ML, Beasley NJ, Ezzat S, Freeman JL, Asa SL. 2002. Overexpression of cyclin D1 and underexpression of p27 predict lymph node metastases in papillary thyroid carcinoma. *J Clin Endocrinol Metab* 87(4):1814-8.
- Kibel AS, Suarez BK, Belani J, Oh J, Webster R, Brophy-Ebbers M, Guo C, Catalona WJ, Picus J, Goodfellow PJ. 2003. CDKN1A and CDKN1B polymorphisms and risk of advanced prostate carcinoma. *Cancer Res* 63(9):2033-6.
- Kirschner LS, Carney JA, Pack SD, Taymans SE, Giatzakis C, Cho YS, Cho-Chung YS, Stratakis CA. 2000. Mutations of the gene encoding the protein kinase A type I-alpha regulatory subunit in patients with the Carney complex. *Nat Genet* 26(1):89-92.
- Kondo T, Ezzat S, Asa SL. 2006. Pathogenetic mechanisms in thyroid follicular-cell neoplasia. *Nat Rev Cancer* 6(4):292-306.
- Kong M, Kim Y, Lee C. 2011. A strong synergistic epistasis between FAM134B and TNFRSF19 on the susceptibility to vascular dementia. *Psychiatr Genet* 21(1):37-41.
- Kroll TG, Sarraf P, Pecciarini L, Chen CJ, Mueller E, Spiegelman BM, Fletcher JA. 2000. PAX8-PPARGgamma1 fusion oncogene in human thyroid carcinoma [corrected]. *Science* 289(5483):1357-60.
- Landa I, Montero-Conde C, Malanga D, De Gisi S, Pita G, Leandro-Garcia LJ, Inglada-Perez L, Leton R, De Marco C, Rodriguez-Antona C and others. 2010. Allelic variant at -79 (C>T) in CDKN1B (p27Kip1) confers an increased risk of thyroid cancer and alters mRNA levels. *Endocr Relat Cancer* 17(2):317-28.
- Landa I, Robledo M. 2011. Association studies in thyroid cancer susceptibility. Are we on the right track? *J Mol Endocrinol* [in press].
- Landa I, Ruiz-Llorente S, Montero-Conde C, Inglada-Perez L, Schiavi F, Leskela S, Pita G, Milne R, Maravall J, Ramos I and others. 2009. The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. *PLoS Genet* 5(9):e1000637.
- Lander ES. 1996. The new genomics: global views of biology. *Science* 274(5287):536-9.
- Ledo F, Carrion AM, Link WA, Mellstrom B, Naranjo JR. 2000. DREAM-alphaCREM interaction via leucine-charged domains derepresses downstream regulatory element-dependent transcription. *Mol Cell Biol* 20(24):9120-6.
- Ledo F, Kremer L, Mellstrom B, Naranjo JR. 2002. Ca²⁺-dependent block of CREB-CBP transcription by repressor DREAM. *Embo J* 21(17):4583-92.
- Lee PH, Shatkay H. 2008. F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic Acids Res* 36(Database issue):D820-4.

- Lee S, Hong SW, Shin SJ, Kim YM, Rhee Y, Jeon BI, Moon WC, Oh MR, Lim SK. 2004. Papillary thyroid carcinoma associated with familial adenomatous polyposis: molecular analysis of pathogenesis in a family and review of the literature. *Endocr J* 51(3):317-23.
- Lemos MC, Carrilho F, Rodrigues F, Coutinho E, Gomes L, Carnevalheiro M, Regateiro FJ. 2007. Genetic polymorphism of CYP2D6 influences susceptibility to papillary thyroid cancer. *Clin Endocrinol (Oxf)* 67(2):180-3.
- Lemos MC, Coutinho E, Gomes L, Carrilho F, Rodrigues F, Regateiro FJ, Carnevalheiro M. 2008. Combined GSTM1 and GSTT1 null genotypes are associated with a lower risk of papillary thyroid cancer. *J Endocrinol Invest* 31(6):542-5.
- Lesueur F, Corbex M, McKay JD, Lima J, Soares P, Griseri P, Burgess J, Ceccherini I, Landolfi S, Papotti M and others. 2002. Specific haplotypes of the RET proto-oncogene are over-represented in patients with sporadic papillary thyroid carcinoma. *J Med Genet* 39(4):260-5.
- Li G, Sturgis EM, Wang LE, Chamberlain RM, Spitz MR, El-Naggar AK, Hong WK, Wei Q. 2004. Association between the V109G polymorphism of the p27 gene and the risk and progression of oral squamous cell carcinoma. *Clin Cancer Res* 10(12 Pt 1):3996-4002.
- Liu J, Singh B, Tallini G, Carlson DL, Katabi N, Shaha A, Tuttle RM, Ghossein RA. 2006. Follicular variant of papillary thyroid carcinoma: a clinicopathologic study of a problematic entity. *Cancer* 107(6):1255-64.
- Liu S, Semenciw R, Ugnat AM, Mao Y. 2001. Increasing thyroid cancer incidence in Canada, 1970-1996: time trends and age-period-cohort effects. *Br J Cancer* 85(9):1335-9.
- Liu Z, Dong Z, Han B, Yang Y, Liu Y, Zhang JT. 2005. Regulation of expression by promoters versus internal ribosome entry site in the 5'-untranslated sequence of the human cyclin-dependent kinase inhibitor p27kip1. *Nucleic Acids Res* 33(12):3763-71.
- Liu Z, Hou P, Ji M, Guan H, Studeman K, Jensen K, Vasko V, El-Naggar AK, Xing M. 2008. Highly prevalent genetic alterations in receptor tyrosine kinases and phosphatidylinositol 3-kinase/akt and mitogen-activated protein kinase pathways in anaplastic and follicular thyroid cancers. *J Clin Endocrinol Metab* 93(8):3106-16.
- Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻($\Delta\Delta C_T$) Method. *Methods* 25(4):402-8.
- Lonn S, Bhatti P, Alexander BH, Pineda MA, Doody MM, Struwing JP, Sigurdson AJ. 2007. Papillary thyroid cancer and polymorphic variants in TSHR- and RET-related genes: a nested case-control study within a cohort of U.S. radiologic technologists. *Cancer Epidemiol Biomarkers Prev* 16(1):174-7.
- Macchia PE, Lapi P, Krude H, Pirro MT, Missero C, Chiovato L, Souabni A, Baserga M, Tassi V, Pinchera A and others. 1998. PAX8 mutations associated with congenital hypothyroidism caused by thyroid dysgenesis. *Nat Genet* 19(1):83-6.
- Malchoff CD, Sarfarazi M, Tendler B, Forouhar F, Whalen G, Joshi V, Arnold A, Malchoff DM. 2000. Papillary thyroid carcinoma associated with papillary renal neoplasia: genetic linkage analysis of a distinct heritable tumor syndrome. *J Clin Endocrinol Metab* 85(5):1758-64.
- Manenti G, Pilotti S, Re FC, Della Porta G, Pierotti MA. 1994. Selective activation of ras oncogenes in follicular and undifferentiated thyroid carcinomas. *Eur J Cancer* 30A(7):987-93.
- Mani SA, Yang J, Brooks M, Schwaninger G, Zhou A, Miura N, Kutok JL, Hartwell K, Richardson AL, Weinberg RA. 2007. Mesenchyme Forkhead 1 (FOXC2) plays a key role in metastasis and is associated with aggressive basal-like breast cancers. *Proc Natl Acad Sci U S A* 104(24):10069-74.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A and others. 2009. Finding the missing heritability of complex diseases. *Nature* 461(7265):747-53.
- Mansouri A, Chowdhury K, Gruss P. 1998. Follicular cells of the thyroid gland require Pax8 gene function. *Nat Genet* 19(1):87-90.
- Matakidou A, Hamel N, Popat S, Henderson K, Kantemiroff T, Harmer C, Clarke SE, Houlston RS, Foulkes WD. 2004. Risk of non-medullary thyroid cancer influenced by polymorphic variation in the thyroglobulin gene. *Carcinogenesis* 25(3):369-73.
- Mazzaferri EL, Kloos RT. 2001. Clinical review 128: Current approaches to primary therapy for papillary and follicular thyroid cancer. *J Clin Endocrinol Metab* 86(4):1447-63.

- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9(5):356-69.
- McKay JD, Lesueur F, Jonard L, Pastore A, Williamson J, Hoffman L, Burgess J, Duffield A, Papotti M, Stark M and others. 2001. Localization of a susceptibility gene for familial nonmedullary thyroid carcinoma to chromosome 2q21. *Am J Hum Genet* 69(2):440-6.
- Mendez-Pertuz M, Sanchez-Pacheco A, Aranda A. 2003. The thyroid hormone receptor antagonizes CREB-mediated transcription. *Embo J* 22(12):3102-12.
- Mestre JR, Rivadeneira DE, Mackrell PJ, Duff M, Stapleton PP, Mack-Strong V, Maddali S, Smyth GP, Tanabe T, Daly JM. 2001. Overlapping CRE and E-box promoter elements can independently regulate COX-2 gene transcription in macrophages. *FEBS Lett* 496(2-3):147-51.
- Milne RL, Benitez J. 2008. Current strategies in the search for low penetrance genes in cancer. *Histol Histopathol* 23(4):507-14.
- Millard SS, Vidal A, Markus M, Koff A. 2000. A U-rich element in the 5' untranslated region is necessary for the translation of p27 mRNA. *Mol Cell Biol* 20(16):5947-59.
- Minami S, Ohtani-Fujita N, Igata E, Tamaki T, Sakai T. 1997. Molecular cloning and characterization of the human p27Kip1 gene promoter. *FEBS Lett* 411(1):1-6.
- Montero-Conde C, Martin-Campos JM, Lerma E, Gimenez G, Martinez-Guitarte JL, Combalia N, Montaner D, Matias-Guiu X, Dopazo J, de Leiva A and others. 2008. Molecular profiling related to poor prognosis in thyroid carcinoma. Combining gene expression data and biological information. *Oncogene* 27(11):1554-61.
- Naidu R, Har YC, Taib NA. 2007. P27 V109G Polymorphism is associated with lymph node metastases but not with increased risk of breast cancer. *J Exp Clin Cancer Res* 26(1):133-40.
- Nakamura Y, Nishisho I, Kinzler KW, Vogelstein B, Miyoshi Y, Miki Y, Ando H, Horii A, Nagase H. 1991. Mutations of the adenomatous polyposis coli gene in familial polyposis coli patients and sporadic colorectal tumors. *Princess Takamatsu Symp* 22:285-92.
- Natali PG, Berlingieri MT, Nicotra MR, Fusco A, Santoro E, Bigotti A, Vecchio G. 1995. Transformation of thyroid epithelium is associated with loss of c-kit receptor. *Cancer Res* 55(8):1787-91.
- Naushad SM, Pavani A, Digumarti RR, Gottumukkala SR, Kutala VK. 2010. Epistatic interactions between loci of one-carbon metabolism modulate susceptibility to breast cancer. *Mol Biol Rep*.
- Nelen MR, Padberg GW, Peeters EA, Lin AY, van den Helm B, Frants RR, Coulon V, Goldstein AM, van Reen MM, Easton DF and others. 1996. Localization of the gene for Cowden disease to chromosome 10q22-23. *Nat Genet* 13(1):114-6.
- Nikiforov YE. 2002. RET/PTC rearrangement in thyroid tumors. *Endocr Pathol* 13(1):3-16.
- Nikiforova MN, Lynch RA, Biddinger PW, Alexander EK, Dorn GW, 2nd, Tallini G, Kroll TG, Nikiforov YE. 2003. RAS point mutations and PAX8-PPAR gamma rearrangement in thyroid tumors: evidence for distinct molecular pathways in thyroid follicular carcinoma. *J Clin Endocrinol Metab* 88(5):2318-26.
- Nose V. 2008. Familial non-medullary thyroid carcinoma: an update. *Endocr Pathol* 19(4):226-40.
- Ohno M, Zannini M, Levy O, Carrasco N, di Lauro R. 1999. The paired-domain transcription factor Pax8 binds to the upstream enhancer of the rat sodium/iodide symporter gene and participates in both thyroid-specific and cyclic-AMP-dependent transcription. *Mol Cell Biol* 19(3):2051-60.
- Ortiz L, Zannini M, Di Lauro R, Santisteban P. 1997. Transcriptional control of the forkhead thyroid transcription factor TTF-2 by thyrotropin, insulin, and insulin-like growth factor I. *J Biol Chem* 272(37):23334-9.
- Pacini F, Castagna MG, Brilli L, Pentheroudakis G. 2010. Thyroid cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 21 Suppl 5:v214-9.
- Pal T, Vogl FD, Chappuis PO, Tsang R, Brierley J, Renard H, Sanders K, Kantemiroff T, Bagha S, Goldgar DE and others. 2001. Increased risk for nonmedullary thyroid cancer in the first degree relatives of prevalent cases of nonmedullary thyroid cancer: a hospital-based study. *J Clin Endocrinol Metab* 86(11):5307-12.

- Papini E, Guglielmi R, Bianchini A, Crescenzi A, Taccogna S, Nardi F, Panunzi C, Rinaldi R, Toscano V, Pacella CM. 2002. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. *J Clin Endocrinol Metab* 87(5):1941-6.
- Parlato R, Rosica A, Rodriguez-Mallon A, Affuso A, Postiglione MP, Arra C, Mansouri A, Kimura S, Di Lauro R, De Felice M. 2004. An integrated regulatory network controlling survival and migration in thyroid organogenesis. *Dev Biol* 276(2):464-75.
- Penna-Martinez M, Ramos-Lopez E, Stern J, Hinsch N, Hansmann ML, Selinski I, Grunwald F, Vorlander C, Wahl RA, Bechstein WO and others. 2009. Vitamin D receptor polymorphisms in differentiated thyroid carcinoma. *Thyroid* 19(6):623-8.
- Pisarev MA, Thomasz L, Juvenal GJ. 2009. Role of transforming growth factor beta in the regulation of thyroid function and growth. *Thyroid* 19(8):881-92.
- Pittman AM, Naranjo S, Webb E, Broderick P, Lips EH, van Wezel T, Morreau H, Sullivan K, Fielding S, Twiss P and others. 2009. The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. *Genome Res* 19(6):987-93.
- Prazeres HJ, Rodrigues F, Soares P, Naidenov P, Figueiredo P, Campos B, Lacerda M, Martins TC. 2008. Loss of heterozygosity at 19p13.2 and 2q21 in tumours from familial clusters of non-medullary thyroid carcinoma. *Fam Cancer* 7(2):141-9.
- Rabes HM, Demidchik EP, Sidorow JD, Lengfelder E, Beimfohr C, Hoelzel D, Klugbauer S. 2000. Pattern of radiation-induced RET and NTRK1 rearrangements in 191 post-chernobyl papillary thyroid carcinomas: biological, phenotypic, and clinical implications. *Clin Cancer Res* 6(3):1093-103.
- Rada-Iglesias A, Ameer A, Kapranov P, Enroth S, Komorowski J, Gingeras TR, Wadelius C. 2008. Whole-genome maps of USF1 and USF2 binding and histone H3 acetylation reveal new aspects of promoter structure and candidate genes for common human disorders. *Genome Res* 18(3):380-92.
- Randolph GW, Maniar D. 2000. Medullary carcinoma of the thyroid. *Cancer Control* 7(3):253-61.
- Riesco-Eizaguirre G, Santisteban P. 2007. Molecular biology of thyroid cancer initiation. *Clin Transl Oncol* 9(11):686-93.
- Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. 2001. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* 69(1):138-47.
- Rivas M, Mellstrom B, Naranjo JR, Santisteban P. 2004. Transcriptional repressor DREAM interacts with thyroid transcription factor-1 and regulates thyroglobulin gene expression. *J Biol Chem* 279(32):33114-22.
- Robbins SL, Cotran RS, Kumar V. 1988. *Patología estructural y funcional*. México [etc.]: : Ineramericana.
- Rodriguez S, Khabir A, Keryer C, Perrot C, Drira M, Ghorbel A, Jlidi R, Bernheim A, Valent A, Busson P. 2005. Conventional and array-based comparative genomic hybridization analysis of nasopharyngeal carcinomas from the Mediterranean area. *Cancer Genet Cytogenet* 157(2):140-7.
- Rogounovitch TI, Saenko VA, Ashizawa K, Sedliarou IA, Namba H, Abrosimov AY, Lushnikov EF, Roumiantsev PO, Konova MV, Petoukhova NS and others. 2006. TP53 codon 72 polymorphism in radiation-associated human papillary thyroid cancer. *Oncol Rep* 15(4):949-56.
- Rueff J, Gaspar J, Kranendonk M. 2002. DNA polymorphisms as modulators of genotoxicity and cancer. *Biol Chem* 383(6):923-32.
- Runkel L, Shaw PE, Herrera RE, Hipskind RA, Nordheim A. 1991. Multiple basal promoter elements determine the level of human c-fos transcription. *Mol Cell Biol* 11(3):1270-80.
- Salajegheh A, Petcu EB, Smith RA, Lam AK. 2008. Follicular variant of papillary thyroid carcinoma: a diagnostic challenge for clinicians and pathologists. *Postgrad Med J* 84(988):78-82.
- Santoro M, Carlomagno F, Hay ID, Herrmann MA, Grieco M, Melillo R, Pierotti MA, Bongarzone I, Della Porta G, Berger N and others. 1992. Ret oncogene activation in human thyroid neoplasms is restricted to the papillary cancer subtype. *J Clin Invest* 89(5):1517-22.
- Santoro M, Dathan NA, Berlingieri MT, Bongarzone I, Paulin C, Grieco M, Pierotti MA, Vecchio G, Fusco A. 1994. Molecular characterization of RET/PTC3; a novel rearranged version of

- the RET proto-oncogene in a human thyroid papillary carcinoma. *Oncogene* 9(2):509-16.
- Schlumberger MJ. 1998. Papillary and follicular thyroid carcinoma. *N Engl J Med* 338(5):297-306.
- Scholtz B, Kingsley-Kallesen M, Rizzino A. 1996. Transcription of the transforming growth factor-beta2 gene is dependent on an E-box located between an essential cAMP response element/activating transcription factor motif and the TATA box of the gene. *J Biol Chem* 271(50):32375-80.
- Schweppe RE, Klopper JP, Korch C, Pugazhenth U, Benezra M, Knauf JA, Fagin JA, Marlow LA, Copland JA, Smallridge RC and others. 2008. Deoxyribonucleic acid profiling analysis of 40 human thyroid cancer cell lines reveals cross-contamination resulting in cell line redundancy and misidentification. *J Clin Endocrinol Metab* 93(11):4331-41.
- Sequeira MJ, Morgan JM, Fuhrer D, Wheeler MH, Jasani B, Ludgate M. 2001. Thyroid transcription factor-2 gene expression in benign and malignant thyroid lesions. *Thyroid* 11(11):995-1001.
- Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D, Yeakley J, Bibikova M, Wickham Garcia E, McBride C and others. 2005. High-throughput SNP genotyping on universal bead arrays. *Mutat Res* 573(1-2):70-82.
- Silva SN, Gil OM, Oliveira VC, Cabral MN, Azevedo AP, Faber A, Manita I, Ferreira TC, Limbert E, Pina JE and others. 2005. Association of polymorphisms in ERCC2 gene with non-familial thyroid cancer risk. *Cancer Epidemiol Biomarkers Prev* 14(10):2407-12.
- Singh B, Gogineni SK, Sacks PG, Shaha AR, Shah JP, Stoffel A, Rao PH. 2001. Molecular cytogenetic characterization of head and neck squamous cell carcinoma and refinement of 3q amplification. *Cancer Res* 61(11):4506-13.
- Siraj AK, Al-Rasheed M, Ibrahim M, Siddiqui K, Al-Dayel F, Al-Sanea O, Uddin S, Al-Kuraya K. 2008a. RAD52 polymorphisms contribute to the development of papillary thyroid cancer susceptibility in Middle Eastern population. *J Endocrinol Invest* 31(10):893-9.
- Siraj AK, Ibrahim M, Al-Rasheed M, Abubaker J, Bu R, Siddiqui SU, Al-Dayel F, Al-Sanea O, Al-Nuaim A, Uddin S and others. 2008b. Polymorphisms of selected xenobiotic genes contribute to the development of papillary thyroid cancer susceptibility in Middle Eastern population. *BMC Med Genet* 9:61.
- Slingerland J, Pagano M. 2000. Regulation of the cdk inhibitor p27 and its deregulation in cancer. *J Cell Physiol* 183(1):10-7.
- Snitkin ES, Segre D. 2011. Epistatic interaction maps relative to multiple metabolic phenotypes. *PLoS Genet* 7(2):e1001294.
- Sozzi G, Bongarzone I, Miozzo M, Borrello MG, Blutti MG, Pilotti S, Della Porta G, Pierotti MA. 1994. A t(10;17) translocation creates the RET/PTC2 chimeric transforming sequence in papillary thyroid carcinoma. *Genes Chromosomes Cancer* 9(4):244-50.
- Spirin KS, Simpson JF, Takeuchi S, Kawamata N, Miller CW, Koeffler HP. 1996. p27/Kip1 mutation found in breast cancer. *Cancer Res* 56(10):2400-4.
- Spurdle AB, Deans AJ, Duffy D, Goldgar DE, Chen X, Beesley J, Easton DF, Antoniou AC, Peock S, Cook M and others. 2008. No evidence that CDKN1B (p27) polymorphisms modify breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Breast Cancer Res Treat*.
- Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68(4):978-89.
- Sturgis EM, Li G. 2009. Molecular epidemiology of papillary thyroid cancer: in search of common genetic associations. *Thyroid* 19(10):1031-4.
- Sturgis EM, Zhao C, Zheng R, Wei Q. 2005. Radiation response genotype and risk of differentiated thyroid cancer: a case-control analysis. *Laryngoscope* 115(6):938-45.
- Subramaniam MM, Putti TC, Anuar D, Chong PY, Shah N, Salto-Tellez M, Soong R. 2007. Clonal characterization of sporadic cribriform-morular variant of papillary thyroid carcinoma by laser microdissection-based APC mutation analysis. *Am J Clin Pathol* 128(6):994-1001.
- Sugg SL, Ezzat S, Rosen IB, Freeman JL, Asa SL. 1998. Distinct multiple RET/PTC gene rearrangements in multifocal papillary thyroid neoplasia. *J Clin Endocrinol Metab* 83(11):4116-22.
- Suh I, Filetti S, Vriens MR, Guerrero MA, Tumino S, Wong M, Shen WT, Kebebew E, Duh QY, Clark OH. 2009. Distinct loci on chromosome 1q21 and 6q22 predispose to familial

- nonmedullary thyroid cancer: a SNP array-based linkage analysis of 38 families. *Surgery* 146(6):1073-80.
- Takahashi M, Saenko VA, Rogounovitch TI, Kawaguchi T, Drozd VM, Takigawa-Imamura H, Akulevich NM, Ratanajaraya C, Mitsutake N, Takamura N and others. 2010. The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. *Hum Mol Genet* 19(12):2516-23.
- Tallini G, Garcia-Rostan G, Herrero A, Zelterman D, Viale G, Bosari S, Carcangiu ML. 1999. Downregulation of p27KIP1 and Ki67/Mib1 labeling index support the classification of thyroid carcinoma into prognostically relevant categories. *Am J Surg Pathol* 23(6):678-85.
- Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, Barnettson RA, Theodoratou E, Cetnarskyj R, Cartwright N and others. 2008. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* 40(5):631-7.
- Tigli H, Buyru N, Dalay N. 2005. Molecular analysis of the p27/kip1 gene in breast cancer. *Mol Diagn* 9(1):17-21.
- Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, Spain S, Lubbe S, Walther A, Sullivan K and others. 2008. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet* 40(5):623-30.
- Viglietto G, Motti ML, Fusco A. 2002. Understanding p27(kip1) deregulation in cancer: down-regulation or mislocalization. *Cell Cycle* 1(6):394-400.
- Viollet B, Lefrancois-Martinez AM, Henrion A, Kahn A, Raymondjean M, Martinez A. 1996. Immunochemical characterization and transacting properties of upstream stimulatory factor isoforms. *J Biol Chem* 271(3):1405-15.
- Vitagliano D, Carlomagno F, Motti ML, Viglietto G, Nikiforov YE, Nikiforova MN, Hershman JM, Ryan AJ, Fusco A, Melillo RM and others. 2004. Regulation of p27Kip1 protein levels contributes to mitogenic effects of the RET/PTC kinase in thyroid carcinoma cells. *Cancer Res* 64(11):3823-9.
- Vriens MR, Suh I, Moses W, Kebebew E. 2009. Clinical features and genetic predisposition to hereditary nonmedullary thyroid cancer. *Thyroid* 19(12):1343-9.
- Wang W, Spitz MR, Yang H, Lu C, Stewart DJ, Wu X. 2007a. Genetic variants in cell cycle control pathway confer susceptibility to lung cancer. *Clin Cancer Res* 13(19):5974-81.
- Wang Y, Hou P, Yu H, Wang W, Ji M, Zhao S, Yan S, Sun X, Liu D, Shi B and others. 2007b. High prevalence and mutual exclusivity of genetic alterations in the phosphatidylinositol-3-kinase/akt pathway in thyroid tumors. *J Clin Endocrinol Metab* 92(6):2387-90.
- Williams D. 2009. Radiation carcinogenesis: lessons from Chernobyl. *Oncogene* 27 Suppl 2:S9-18.
- Wlodarska I, Veyt E, De Paepe P, Vandenberghe P, Nooijen P, Theate I, Michaux L, Sagaert X, Marynen P, Hagemeijer A and others. 2005. FOXP1, a gene highly expressed in a subset of diffuse large B-cell lymphoma, is recurrently targeted by genomic aberrations. *Leukemia* 19(8):1299-305.
- Wokolorczyk D, Gliniewicz B, Sikorski A, Zlowocka E, Masojc B, Debniak T, Matyjasik J, Mierzejewski M, Medrek K, Oszutowska D and others. 2008. A range of cancers is associated with the rs6983267 marker on chromosome 8. *Cancer Res* 68(23):9982-6.
- Woodruff SL, Arowolo OA, Akute OO, Afolabi AO, Nwariaku F. 2010. Global variation in the pattern of differentiated thyroid cancer. *Am J Surg* 200(4):462-6.
- Wu SH, Lee KW, Chen CH, Lin CC, Tseng YM, Ma H, Tsai SM, Tsai LY. 2010. Epistasis of oxidative stress-related enzyme genes on modulating the risks in oral cavity cancer. *Clin Chim Acta* 411(21-22):1705-10.
- Xing M. 2005. BRAF mutation in thyroid cancer. *Endocr Relat Cancer* 12(2):245-62.
- Young B, Heath JW, Stevens A, Lowe JS, Wheeler PR, Burkitt HG. 2000. Wheeler's Functional histology : a text and colour atlas. Edinburgh ; New York: Churchill Livingstone.
- Zannini M, Francis-Lang H, Plachov D, Di Lauro R. 1992. Pax-8, a paired domain-containing protein, binds to a sequence overlapping the recognition site of a homeodomain and activates transcription from two thyroid-specific promoters. *Mol Cell Biol* 12(9):4230-41.

- Zeng YX, Somasundaram K, el-Deiry WS. 1997. AP2 inhibits cancer cell growth and activates p21WAF1/CIP1 expression. *Nat Genet* 15(1):78-82.
- Zhu Z, Gandhi M, Nikiforova MN, Fischer AH, Nikiforov YE. 2003. Molecular profile and clinical-pathologic features of the follicular variant of papillary thyroid carcinoma. An unusually high prevalence of ras mutations. *Am J Clin Pathol* 120(1):71-7.
- Zidan J, Karen D, Stein M, Rosenblatt E, Basher W, Kuten A. 2003. Pure versus follicular variant of papillary thyroid carcinoma: clinical features, prognostic factors, treatment, and survival. *Cancer* 97(5):1181-5.

APÉNDICE I

Material suplementario

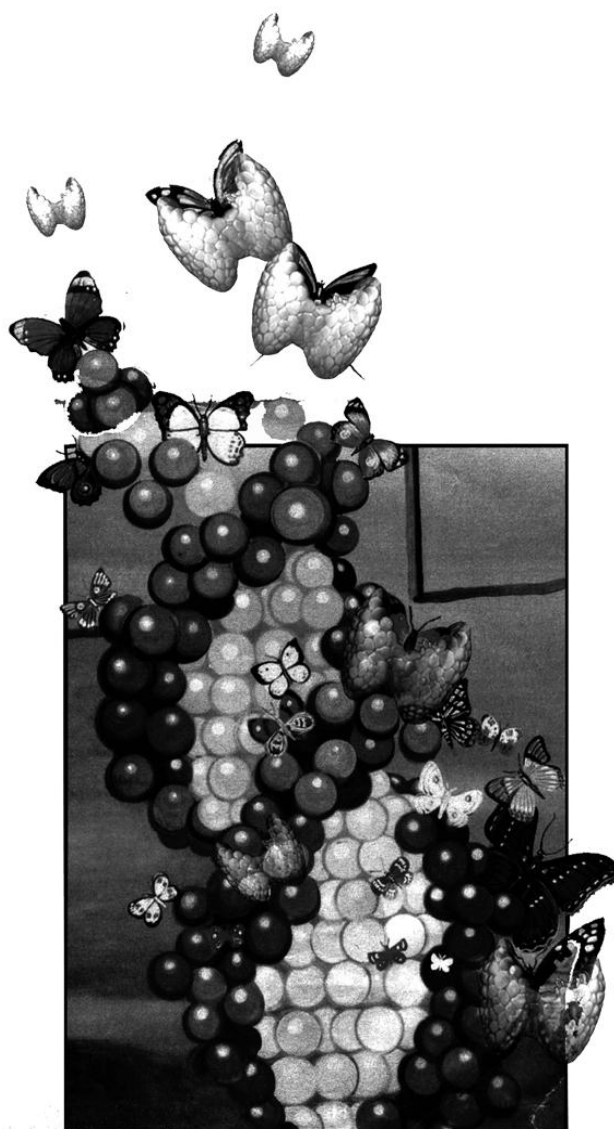


Tabla S1. Listado de SNPs estudiados en la serie 1 del estudio de asociación.

1_23935574_rs10917420_tag_TCEB3_b1	2_1499683_rs1126799_no.sin_TPO_r1
1_23944942_rs550850_tag_TCEB3_b1	2_1519491_rs10181280_tag_TPO_b13
1_23948463_rs2294495_tag_TCEB3_b1	2_1524117_rs4927632_tag_TPO_b13
1_23950038_rs2235541_no.sin_TCEB3_b1	2_1525334_rs1042589_no.sin_TPO_b13
1_23956236_rs2076346_tag_TCEB3_b1	2_43400339_rs9808360_tag_THADA_b5
1_54138952_rs12131461_tag_DIO1_b1	2_43401176_rs17334247_no.sin_THADA_b5
1_54148289_rs11206244_ESE_DIO1_r1	2_43401632_rs17030684_tag_THADA_b5
1_54153994_rs12063777_tag_DIO1_b2	2_43407453_rs1465618_tag_THADA_b5
1_65075926_rs310241_tag_JAK1_b3	2_43408233_rs4128209_tag_THADA_b5
1_65076098_rs4916004_tag_JAK1_b3	2_43410988_rs17406174_tag_THADA_b5
1_65076247_rs12129819_tag_JAK1_b3	2_43469512_rs6739828_tag_THADA_b9
1_65077058_rs11579616_tag_JAK1_b3	2_43472323_rs11899863_tag_THADA_b9
1_65078465_rs310244_tag_JAK1_b3	2_43472544_rs10166054_tag_THADA_b9
1_65079997_rs310247_tag_JAK1_b3	2_43478582_rs7605661_tag_THADA_b9
1_65083077_rs2230588_ESE_JAK1_b4	2_43481559_rs7600177_tag_THADA_b9
1_65083802_rs3737139_tag_JAK1_b4	2_43483726_rs17030749_tag_THADA_b9
1_65083850_rs2230587_tag_JAK1_b4	2_43483963_rs17030751_TFBS_THADA_b9
1_65084930_rs17127063_ESE_JAK1_b4	2_43530443_rs12105399_tag_THADA_b16
1_65093838_rs11579263_ESE_JAK1_b4	2_43539241_rs6738365_tag_THADA_b16
1_65106445_rs310236_tag_JAK1_b5	2_43539711_rs2046916_tag_THADA_b16
1_65108419_rs17127114_tag_JAK1_b5	2_43562870_rs999949_tag_THADA_b16
1_65115093_rs12135754_tag_JAK1_b5	2_43564570_rs4146253_tag_THADA_b16
1_65118204_rs310209_tag_JAK1_b5	2_43571154_rs6746064_tag_THADA_b16
1_65130866_rs17390852_TFBS_JAK1_r1	2_43586327_rs7578597_no.sin_THADA_b16
1_115366523_rs11102872_TFBS_TSHB_b1	2_43591677_rs1038822_tag_THADA_b16
1_115367714_rs7515454_TFBS_TSHB_b1	2_43608496_rs1491508_tag_THADA_b16
1_115367826_rs7517739_TFBS_TSHB_b1	2_43620543_rs7559891_tag_THADA_b17
1_115368615_rs12040783_tag_TSHB_b2	2_43630357_rs6742108_tag_THADA_b17
1_115369408_rs1575070_TFBS_TSHB_b2	2_43631158_rs11903287_tag_THADA_b17
1_115369551_rs1575069_tag_TSHB_b2	2_43633169_rs1873555_tag_THADA_b17
1_115373439_rs1321109_tag_TSHB_b2	2_43639779_rs17413525_tag_THADA_b17
1_115377238_rs11102875_tag_TSHB_b2	2_43651123_rs17031056_no.sin_THADA_b17
1_115377546_rs10776792_no.sin_TSHB_b2	2_43655216_rs11899823_ESE_THADA_b17
1_115382063_rs1998008_tag_TSHB_b2	2_43683044_rs10182489_tag_THADA_b17
1_115387924_rs4611011_tag_TSHB_b2	2_95319407_rs2278067_tag_KCNIP3_b1
1_149056488_rs10305741_no.sin_ARNT_r1	2_95323539_rs889854_tag_KCNIP3_b1
1_150212253_rs3007700_tag_S100A10_b1	2_95326275_rs3755525_tag_KCNIP3_b1
1_150212403_rs16833728_tag_S100A10_b1	2_95329773_rs3772044_tag_KCNIP3_b1
1_150218187_rs12128371_tag_S100A10_b1	2_95352576_rs2113418_tag_KCNIP3_b1
1_150218319_rs2999526_tag_S100A10_b1	2_95387358_rs3755518_tag_KCNIP3_b1
1_150221633_rs6678672_tag_S100A10_b1	2_95387391_rs3821340_tag_KCNIP3_b1
1_150228199_rs11204922_tag_S100A10_b1	2_95398530_rs6712299_tag_KCNIP3_b3
1_150233369_rs1038745_TFBS_S100A10_b1	2_95420738_rs2320433_tag_KCNIP3_b3
1_155041296_rs2644609_tag_NTRK1_b1	2_113690667_rs874898_tag_PAX8_b2
1_155043341_rs928391_tag_NTRK1_b1	2_113691537_rs1478_ESE_PAX8_b2
1_155052241_rs1800601_tag_NTRK1_b2	2_113695411_rs11123170_tag_PAX8_b3
1_155053552_rs2768764_tag_NTRK1_b2	2_113700974_rs2241975_tag_PAX8_b3
1_155059756_rs7517728_tag_NTRK1_b3	2_113705707_rs2863242_tag_PAX8_b3
1_155061630_rs4661222_tag_NTRK1_b3	2_113717336_rs2863244_tag_PAX8_b4
1_155063468_rs6427332_tag_NTRK1_b3	2_113718391_rs13007173_tag_PAX8_b4
1_155069281_rs4661061_tag_NTRK1_b3	2_113720374_rs1491584_tag_PAX8_b4
1_155075830_rs4661229_tag_NTRK1_b3	2_113733499_rs6734610_tag_PAX8_b5
1_155091425_rs4661063_TFBS_NTRK1_r1	2_113735060_rs895417_tag_PAX8_b5
1_155101171_rs6334_ESE_NTRK1_r2	2_113737318_rs1466018_tag_PAX8_b5
1_155112857_rs6334_ESE_NTRK1_r3	2_113739039_rs6542126_tag_PAX8_b5
1_155115542_rs6336_no.sin_NTRK1_r4	2_113741872_rs1446246_tag_PAX8_b5
1_155115619_rs6337_ESE_NTRK1_r5	2_113749237_rs4848323_tag_PAX8_b5
1_234448448_rs1055851_no.sin_ERO1LB_r1	2_113758446_rs17626434_tag_PAX8_b6
2_1388360_rs9677074_tag_TPO_b1	2_113763028_rs11123176_tag_PAX8_b6
2_1389936_rs17786733_TFBS_TPO_b1	2_136584543_rs13022389_tag_CXCR4_b2
2_1393966_rs2298873_TFBS_TPO_b1	2_136586343_rs2734871_tag_CXCR4_b2
2_1470181_rs7574221_tag_TPO_b10	2_136589554_rs2228014_ESE_CXCR4_b2
2_1475429_rs11687944_tag_TPO_b10	2_136595086_rs9973445_tag_CXCR4_b2
2_1478934_rs732609_no.sin_TPO_b10	2_136596944_rs12691874_tag_CXCR4_b2
2_1481670_rs3755552_tag_TPO_b10	2_136598575_rs6751768_tag_CXCR4_b2
2_1487692_rs4927621_tag_TPO_b10	2_160516321_rs3828323_no.sin_PLA2R1_r1
2_1493291_rs13414566_tag_TPO_b10	2_160593688_rs3749117_no.sin_PLA2R1_r2
	2_196701858_rs12469216_tag_STK17B_b1
	2_196709152_rs7581626_tag_STK17B_b1
	2_196719027_rs16845711_tag_STK17B_b1
	2_196753322_rs1378624_tag_STK17B_b1
	2_215944957_rs17449032_no.sin_FN1_r1

2_227368788_rs1801278_no.sin_IRS1_r1	4_55283293_rs2298976_tag_KIT_b5
2_227369287_rs1801123_ESE_IRS1_r2	4_55283476_rs11735550_tag_KIT_b5
2_227370997_rs2234931_ESE_IRS1_r3	4_55285378_rs4864920_tag_KIT_b5
2_230324789_rs7565973_tag_TRIP12_b1	4_55294193_rs1008658_tag_KIT_b7
2_230331371_rs7571826_tag_TRIP12_b1	4_55294413_rs1124007_tag_KIT_b7
2_230333546_rs6758201_tag_TRIP12_b1	4_55296940_rs12643468_tag_KIT_b7
2_230337382_rs1044822_tag_TRIP12_b1	4_55297522_rs3733542_tag_KIT_b7
2_230338935_rs11696097_tag_TRIP12_b1	4_55299697_rs17084733_ESE_KIT_b7
2_230340512_rs6687_ESE_TRIP12_b1	4_55301967_rs3796776_tag_KIT_b7
2_230393214_rs488007_tag_TRIP12_b1	4_55304790_rs4356963_tag_KIT_b7
2_230432021_rs544480_ESE_TRIP12_b1	4_141702921_rs2270565_no.sin_UCP_r1
2_230472470_rs17258702_tag_TRIP12_b1	5_42659309_rs6887528_tag_GHR_b5
2_230479357_rs831372_tag_TRIP12_b1	5_42659702_rs7727694_tag_GHR_b5
2_230495305_rs12993164_TFBS_TRIP12_b1	5_42667021_rs6873545_tag_GHR_b5
3_4691811_rs2306877_ESE_ITPR1_r1	5_42715853_rs4547964_tag_GHR_b5
3_4817231_rs2291862_ESE_ITPR1_r2	5_42735801_rs6179_ESE_GHR_b5
3_4831180_rs711631_ESE_ITPR1_r3	5_42753367_rs12187996_tag_GHR_b5
3_4831234_rs901854_ESE_ITPR1_r4	5_42754996_rs6180_no.sin_GHR_b5
3_12309416_rs2972164_tag_PPARG_b1	5_42755583_rs2910875_ESE_GHR_b5
3_12319441_rs7620165_tag_PPARG_b1	5_42760670_rs1364024_tag_GHR_b5
3_12334049_rs2067819_tag_PPARG_b1	5_52237479_rs4145748_no.sin_ITGA_r1
3_12347386_rs4145573_tag_PPARG_b1	5_52250338_rs2279587_no.sin_ITGA_r2
3_12363563_rs10510418_tag_PPARG_b1	5_52265502_rs12520591_no.sin_ITGA_r3
3_12368125_rs1801282_no.sin_PPARG_b1	5_55259686_rs13357543_tag_IL6ST_b1
3_12377474_rs1373641_tag_PPARG_b2	5_55272131_rs10471960_tag_IL6ST_b1
3_12388339_rs2120825_tag_PPARG_b2	5_55275763_rs1900173_tag_IL6ST_b1
3_12389420_rs6809832_tag_PPARG_b2	5_55279795_rs11574780_tag_IL6ST_b1
3_12399793_rs2972162_tag_PPARG_b2	5_55280534_rs7719246_tag_IL6ST_b1
3_12412237_rs4135268_tag_PPARG_b2	5_55287688_rs2228043_no.sin_IL6ST_b1
3_12418844_rs4135275_tag_PPARG_b2	5_55298417_rs10940495_tag_IL6ST_b1
3_12437024_rs709157_tag_PPARG_b2	5_55336283_rs6450361_TFBS_IL6ST_r1
3_12441214_rs1175542_tag_PPARG_b2	5_80204693_rs26279_no.sin_MSH3_r2
3_12443589_rs13099828_tag_PPARG_b3	5_89979327_rs2366777_no.sin_MASS1_r1
3_12445239_rs1797912_tag_PPARG_b3	5_90015274_rs17544552_no.sin_MASS1_r2
3_12450088_rs7626560_tag_PPARG_b3	5_90015506_rs16868972_no.sin_MASS1_r4
3_12450557_rs3856806_ESE_PPARG_r1	5_90021638_rs10037067_no.sin_MASS1_r5
3_12599070_rs15997_tag_RAF1_b1	5_90024260_rs2366926_no.sin_MASS1_r6
3_12599763_rs7956_tag_RAF1_b1	5_90035966_rs16869016_no.sin_MASS1_r8
3_12600747_rs1051208_ESE_RAF1_b1	5_90052627_rs16869032_no.sin_MASS1_r9
3_12601394_rs3730296_ESE_RAF1_b1	5_90142864_rs2438374_no.sin_MASS1_r11
3_12601516_rs3729931_tag_RAF1_b1	5_90187345_rs2247870_no.sin_MASS1_r12
3_12619335_rs1039244_tag_RAF1_b2	5_131429914_rs11575022_TFBS_CSF2_r1
3_12621255_rs6777175_TFBS_RAF1_b2	5_131430637_rs3091338_tag_CSF2_b1
3_12631016_rs4234512_tag_RAF1_b2	5_131431992_rs11242103_tag_CSF2_b1
3_12634494_rs4684867_tag_RAF1_b2	5_131432345_rs31473_tag_CSF2_b1
3_12637370_rs11705805_tag_RAF1_b2	5_131434943_rs25879_tag_CSF2_b1
3_12638835_rs11923427_tag_RAF1_b2	5_131435392_rs4705916_TFBS_CSF2_b1
3_12680048_rs5746154_tag_RAF1_b2	5_131439037_rs25881_tag_CSF2_b1
3_37028572_rs1799977_no.sin_MLH_r1	5_131439359_rs25882_no.sin_CSF2_b2
3_46900543_rs11926707_tag_PTHR1_b1	5_131441154_rs27438_tag_CSF2_b2
3_46904550_rs6442037_tag_PTHR1_b1	5_176445853_rs640180_tag_FGFR4_b1
3_46906482_rs4683301_tag_PTHR1_b2	5_176450067_rs446382_ESE_FGFR4_b1
3_46910531_rs724450_tag_PTHR1_b2	5_176450403_rs376618_no.sin_FGFR4_b1
3_46916120_rs2242116_tag_PTHR1_b2	5_176452849_rs351855_no.sin_FGFR4_b1
3_46923530_rs7652849_tag_PTHR1_b2	5_176460183_rs6556301_tag_FGFR4_b1
3_46926939_rs11719795_tag_PTHR1_b2	5_176464378_rs7708357_tag_FGFR4_b1
3_46930408_rs4682845_tag_PTHR1_b2	6_4661015_rs3812179_no.sin_CDYL_r1
3_151941695_rs8403_ESE_SIAH2_r1	6_33767450_rs2229642_no.sin_ITPR3_r1
3_151970635_rs2014279_TFBS_SIAH2_r2	6_87845883_rs17449571_tag_CGA_b1
4_55216917_rs6554198_tag_KIT_b2	6_87850120_rs9362387_tag_CGA_b1
4_55219490_rs3822213_tag_KIT_b2	6_87851299_rs9359730_tag_CGA_b1
4_55220687_rs2237037_tag_KIT_b2	6_87851978_rs6631_tag_CGA_b1
4_55221008_rs2237035_tag_KIT_b2	6_87852029_rs1055409_ESE_CGA_b2
4_55223410_rs2237034_tag_KIT_b2	6_87852582_rs7745823_tag_CGA_b2
4_55226632_rs2237032_tag_KIT_b2	6_87854623_rs6155_ESE_CGA_b2
4_55230093_rs2237029_tag_KIT_b3	6_87860673_rs6924434_tag_CGA_b2
4_55231132_rs2237028_tag_KIT_b3	6_87862004_rs9344675_TFBS_CGA_b2
4_55232238_rs759083_tag_KIT_b3	6_87865149_rs1998616_TFBS_CGA_b2
4_55236636_rs2237025_tag_KIT_b4	6_87868159_rs17450398_TFBS_CGA_b2
4_55247087_rs13135792_tag_KIT_b4	6_87870842_rs6909901_tag_CGA_b3
4_55249088_rs2703488_tag_KIT_b4	6_87870921_rs2025488_TFBS_CGA_b3
4_55266198_rs3111800_tag_KIT_b4	7_5993133_rs1805318_no.sin_PMS2_r1

7_43589331_rs10251937_ESE_STK17A_b1	8_104444492_rs827595_tag_CTHRC1_b1
7_43602694_rs10271892_tag_STK17A_b1	8_104445801_rs827598_TFBS_CTHRC1_b1
7_43611958_rs7802995_tag_STK17A_b1	8_104446089_rs827599_tag_CTHRC1_b1
7_43620734_rs10259269_tag_STK17A_b2	8_104454892_rs2575695_tag_CTHRC1_b1
7_43624282_rs12702053_tag_STK17A_b2	8_104463920_rs3098233_ESE_CTHRC1_b1
7_43625329_rs4077509_tag_STK17A_b3	8_120033233_rs2073618_no.sin_TNFRSF11B_r1
7_43625707_rs6950861_tag_STK17A_b3	8_133938444_rs1916273_tag_TG_b1
7_43626289_rs4077435_tag_STK17A_b3	8_133939006_rs2467993_TFBS_TG_b1
7_43627418_rs1053080_tag_STK17A_b3	8_133946112_rs17683567_tag_TG_b1
7_43629948_rs3779062_no.sin_STK17A_b3	8_133946609_rs180194_TFBS_TG_b1
7_43630805_rs1044141_no.sin_STK17A_b3	8_133949394_rs180202_tag_TG_b1
7_43631002_rs15866_tag_STK17A_b3	8_133950426_rs180204_tag_TG_b1
7_43631310_rs1044217_ESE_STK17A_b3	8_133968342_rs180222_no.sin_TG_b1
7_43637757_rs2330918_tag_STK17A_b3	8_133969434_rs180223_no.sin_TG_b1
7_45899194_rs4619_no.sin_IGFBP_r1	8_133969568_rs2069550_ESE_TG_b1
7_100328733_rs1799805_no.sin_ACHE_r1	8_133975788_rs10435562_tag_TG_b1
7_100329387_rs17885778_no.sin_ACHE_r2	8_133979156_rs853326_no.sin_TG_b1
7_107081387_rs2712230_tag_SLC26A4_b1	8_133986770_rs2068128_tag_TG_b2
7_107086763_rs2701684_tag_SLC26A4_b1	8_133989700_rs2069556_no.sin_TG_b2
7_107087576_rs2712228_tag_SLC26A4_b1	8_133991009_rs2687834_tag_TG_b2
7_107090864_rs2248465_tag_SLC26A4_b1	8_133995875_rs2687820_tag_TG_b2
7_107095577_rs6949189_tag_SLC26A4_b2	8_134000365_rs853305_tag_TG_b2
7_107101183_rs11769313_tag_SLC26A4_b2	8_134000930_rs853304_ESE_TG_b2
7_107104411_rs10276013_tag_SLC26A4_b2	8_134028366_rs2205452_tag_TG_b2
7_107108438_rs10250105_tag_SLC26A4_b2	8_134041895_rs2142384_tag_TG_b3
7_107115551_rs1858929_tag_SLC26A4_b4	8_134044465_rs2069561_no.sin_TG_b3
7_107121262_rs2237679_tag_SLC26A4_b5	8_134050928_rs17693031_ESE_TG_r1
7_107137863_rs17154353_no.sin_SLC26A4_b5	8_134052230_rs10106773_tag_TG_b4
7_107146620_rs12667481_tag_SLC26A4_b5	8_134053240_rs11535853_no.sin_TG_b4
7_107153087_rs11972418_tag_SLC26A4_b5	8_134054720_rs7015412_tag_TG_b4
7_128633564_rs2228617_tag_SMO_b1	8_134062846_rs3892189_tag_TG_b4
7_128638111_rs2016607_ESE_SMO_b1	8_134177081_rs3779955_tag_TG_b13
7_128639617_rs2075780_ESE_SMO_b1	8_134177635_rs2069568_ESE_TG_b13
7_128640089_rs3824_ESE_SMO_b1	8_134191878_rs2979033_tag_TG_b15
7_128640256_rs1061275_ESE_SMO_b1	8_134192567_rs4736435_tag_TG_b15
7_128640428_rs1061282_ESE_SMO_b1	8_134194096_rs2958681_tag_TG_b15
7_128640544_rs1061285_ESE_SMO_b1	8_134194211_rs2979040_tag_TG_b15
7_155297702_rs9333594_ESE_SHH_r1	8_134194864_rs1133076_no.sin_TG_b15
7_155299433_rs288746_tag_SHH_b1	8_134199608_rs10096431_tag_TG_b15
7_155305818_rs6948512_tag_SHH_b1	8_134209728_rs17705719_tag_TG_b17
7_155308388_rs172310_tag_SHH_b1	8_134212816_rs2979051_tag_TG_b17
8_19307283_rs3802330_ESE_ChGn_r1	8_134213092_rs17633144_tag_TG_b17
8_19389160_rs11204047_tag_ChGn_b9	8_134213295_rs2294024_ESE_TG_b17
8_19396413_rs17479581_tag_ChGn_b9	8_134223056_rs10755938_tag_TG_b17
8_19402576_rs4484715_tag_ChGn_b9	9_4965442_rs4742060_tag_JAK2_b1
8_19407048_rs7017776_no.sin_ChGn_b9	9_4969730_rs1887427_tag_JAK2_b2
8_19407705_rs6984644_ESE_ChGn_b9	9_4971602_rs10758669_tag_JAK2_b2
8_19493906_rs7841402_tag_ChGn_b17	9_4973311_rs3808850_tag_JAK2_b2
8_19500550_rs4922066_tag_ChGn_b17	9_4988639_rs2225125_tag_JAK2_b3
8_19500753_rs4244454_no.sin_ChGn_b17	9_4993338_rs4372063_tag_JAK2_b3
8_19584012_rs6985674_tag_ChGn_b25	9_5000192_rs10815144_tag_JAK2_b3
8_19585246_rs4548199_tag_ChGn_b25	9_5005732_rs7046736_tag_JAK2_b3
8_19587601_rs1994789_tag_ChGn_b25	9_5006145_rs10815147_TFBS_JAK2_b3
8_19588068_rs17481221_TFBS_ChGn_b25	9_5011514_rs7034753_TFBS_JAK2_b3
8_19591563_rs4922086_tag_ChGn_b25	9_5045434_rs1536800_tag_JAK2_b3
8_38405382_rs2288696_tag_FGFR1_b1	9_5049440_rs2149556_tag_JAK2_b3
8_38406370_rs2915665_ESE_FGFR1_b1	9_5062491_rs10974945_ESE_JAK2_b3
8_38408060_rs2978083_tag_FGFR1_b1	9_5062602_rs17490221_no.sin_JAK2_b3
8_38412051_rs2978076_tag_FGFR1_b1	9_5062846_rs10974947_tag_JAK2_b3
8_38418872_rs6987534_tag_FGFR1_b1	9_5071780_rs2230724_ESE_JAK2_b3
8_38441503_rs6996321_tag_FGFR1_b2	9_5073173_rs1410779_tag_JAK2_b3
8_38444952_rs2467531_ESE_FGFR1_b2	9_97247123_rs16909865_ESE_PTCH_r1
8_38451252_rs7829058_tag_FGFR1_b2	9_97249415_rs357564_no.sin_PTCH_r2
8_38451475_rs881301_tag_FGFR1_b2	9_97278179_rs2066836_ESE_PTCH_r3
8_42155734_rs1136159_no.sin_PLAT_r1	9_99645254_rs7048394_tag_FOXE1_b1
8_42297491_rs17875749_no.sin_IKBKB_r1	9_99652091_rs894673_tag_FOXE1_b1
8_42307589_rs17611716_no.sin_IKBKB_r2	9_99653961_rs3758249_TFBS_FOXE1_b1
8_99012774_rs2290472_no.sin_MATN2_r1	9_99654938_rs907577_TFBS_FOXE1_b1
8_99060397_rs1869609_no.sin_MATN2_r2	9_99656842_rs3021526_tag_FOXE1_b1
8_99113704_rs2255317_no.sin_MATN2_r3	9_99661939_rs874004_tag_FOXE1_b1
8_104443583_rs10096207_TFBS_CTHRC1_b1	9_99664423_rs10119760_tag_FOXE1_b1
8_104443839_rs10109642_tag_CTHRC1_b1	10_51226455_rs2072701_tag_NCOA4_b1

10_51228131_rs7076948_tag_NCOA4_b1	11_63813658_rs12794369_TFBS_BAD_r1
10_51228666_rs10994470_tag_NCOA4_b1	11_101899234_rs17886506_no.sin_MMP7_r1
10_51229475_rs7904463_tag_NCOA4_b1	11_101903803_rs10502001_no.sin_MMP7_r2
10_51231805_rs17178655_tag_NCOA4_b1	11_130255852_rs2298566_no.sin_SNX19_r1
10_51240158_rs10740051_tag_NCOA4_b2	11_130281734_rs4414223_no.sin_SNX19_r2
10_51241137_rs3813713_tag_NCOA4_b2	11_130289964_rs3751037_no.sin_SNX19_r3
10_51264468_rs7350420_tag_NCOA4_b2	12_5998704_rs216311_no.sin_VWF_r1
10_61209408_rs721992_tag_CCDC6_b1	12_6014245_rs216321_no.sin_VWF_r2
10_61209787_rs12243559_tag_CCDC6_b1	12_6023795_rs1063856_no.sin_VWF_r3
10_61212482_rs1913512_tag_CCDC6_b1	12_6042463_rs1800378_no.sin_VWF_r4
10_61216015_rs4948364_tag_CCDC6_b1	12_27999841_rs12227702_tag_PTHLH_b1
10_61217339_rs12267686_tag_CCDC6_b1	12_28000415_rs3910902_tag_PTHLH_b1
10_61218808_rs16914105_tag_CCDC6_b1	12_28007324_rs6245_ESE_PTHLH_b1
10_61219824_rs11540401_ESE_CCDC6_b1	12_28007378_rs2796_tag_PTHLH_b1
10_61220063_rs3802695_ESE_CCDC6_b1	12_28010114_rs805512_tag_PTHLH_b2
10_61222698_rs1053266_no.sin_CCDC6_b2	12_28010651_rs997199_tag_PTHLH_b2
10_61230519_rs10994021_tag_CCDC6_b2	12_28021839_rs11049248_tag_PTHLH_b3
10_61230824_rs12571632_tag_CCDC6_b2	12_52151122_rs11170562_tag_MAP3K12_b1
10_61231434_rs4948242_tag_CCDC6_b2	12_52152886_rs784563_tag_MAP3K12_b1
10_61231983_rs7079829_tag_CCDC6_b2	12_52159633_rs4790_tag_MAP3K12_b1
10_61236006_rs10509112_tag_CCDC6_b2	12_52160111_rs1049193_tag_MAP3K12_b1
10_61246856_rs7088886_tag_CCDC6_b3	12_52166920_rs7958457_tag_MAP3K12_b1
10_61250973_rs920964_tag_CCDC6_b3	12_55014267_rs2371494_tag_STAT2_b1
10_61251195_rs17201105_tag_CCDC6_b3	12_55026949_rs2066807_no.sin_STAT2_b1
10_61255837_rs10994034_tag_CCDC6_b3	12_100074307_rs2279834_ESE_SLC5A8_b1
10_61270958_rs4948372_tag_CCDC6_b4	12_100080857_rs7969185_tag_SLC5A8_b1
10_61273589_rs3851247_tag_CCDC6_b4	12_100084459_rs164365_no.sin_SLC5A8_b1
10_61282038_rs17791436_tag_CCDC6_b4	12_100085094_rs164364_tag_SLC5A8_b1
10_61284289_rs1709343_tag_CCDC6_b4	12_100099783_rs2671434_tag_SLC5A8_b1
10_61284917_rs2440926_tag_CCDC6_b4	12_100111649_rs1709189_no.sin_SLC5A8_b1
10_61286390_rs17202345_tag_CCDC6_b4	12_100116776_rs11110696_tag_SLC5A8_b1
10_61286612_rs1664284_tag_CCDC6_b5	12_100118718_rs10860698_tag_SLC5A8_b2
10_61286730_rs1664285_tag_CCDC6_b5	12_100119700_rs17410881_tag_SLC5A8_b2
10_61293808_rs10509114_tag_CCDC6_b6	12_100124000_rs12580634_tag_SLC5A8_b3
10_61293894_rs2458663_tag_CCDC6_b6	12_100124452_rs2712622_tag_SLC5A8_b3
10_61294703_rs1177701_tag_CCDC6_b7	12_100126478_rs7962305_tag_SLC5A8_b3
10_61299829_rs1621496_tag_CCDC6_b7	12_100128353_rs1399236_tag_SLC5A8_b3
10_61300256_rs1684898_tag_CCDC6_b7	12_100129541_rs2625160_tag_SLC5A8_b3
10_61301422_rs1664256_tag_CCDC6_b8	12_100131430_rs350463_TFBS_SLC5A8_b3
10_61301769_rs1664257_tag_CCDC6_b8	12_100135169_rs11110700_tag_SLC5A8_b3
10_61302122_rs1684900_tag_CCDC6_b8	12_100137619_rs2712623_tag_SLC5A8_b4
10_61302634_rs1684901_tag_CCDC6_b8	12_100138534_rs2062167_tag_SLC5A8_b4
10_61314974_rs1171813_tag_CCDC6_b9	12_119901033_rs1169288_no.sin_TCF1_r1
10_61321707_rs1180657_tag_CCDC6_b10	12_119919810_rs2464196_no.sin_TCF1_r2
10_61325303_rs1171812_tag_CCDC6_b10	12_119919858_rs2464195_no.sin_TCF1_r3
10_61326899_rs1125168_tag_CCDC6_b10	12_123387366_rs2230944_no.sin_NCOR2_r1
10_61327506_rs2393573_tag_CCDC6_b10	14_80481870_rs12895801_tag_TSHR_b1
10_61330162_rs3793866_tag_CCDC6_b10	14_80484500_rs10146516_TFBS_TSHR_b1
10_61331570_rs7914022_tag_CCDC6_b10	14_80485906_rs12050279_tag_TSHR_b1
10_61338345_rs3763745_tag_CCDC6_b10	14_80601507_rs11845164_tag_TSHR_b10
10_61338573_rs1171833_tag_CCDC6_b11	14_80620641_rs2300527_tag_TSHR_b10
10_61339057_rs3763743_tag_CCDC6_b11	14_80623486_rs2284734_tag_TSHR_b10
10_75341362_rs2227580_no.sin_PLAU_r1	14_80623539_rs2284735_tag_TSHR_b10
10_75343107_rs2227564_no.sin_PLAU_r2	14_80629997_rs12147434_tag_TSHR_b10
11_3795486_rs7946577_tag_RHOG_b1	14_80632751_rs2075179_ESE_TSHR_b10
11_3795529_rs10835147_tag_RHOG_b1	14_80636833_rs2300534_tag_TSHR_b11
11_3803869_rs1055640_tag_RHOG_b2	14_80638353_rs7143914_tag_TSHR_b11
11_3808016_rs7948532_tag_RHOG_b3	14_80639431_rs1017141_tag_TSHR_b11
11_3809528_rs1451721_tag_RHOG_b3	14_80640841_rs7150670_tag_TSHR_b11
11_3812435_rs4597058_tag_RHOG_b4	14_80641412_rs7151769_tag_TSHR_b11
11_3813129_rs1451722_tag_RHOG_b4	14_80644758_rs3783941_no.sin_TSHR_b11
11_3813244_rs1451724_tag_RHOG_b4	14_80645172_rs2080305_tag_TSHR_b11
11_3815192_rs11030008_tag_RHOG_b4	14_80680336_rs1991517_no.sin_TSHR_b11
11_3823926_rs11030043_tag_RHOG_b5	14_91496879_rs4904828_tag_TRIP11_b1
11_32417040_rs6508_no.sin_WT1_r1	14_91496952_rs7154514_tag_TRIP11_b1
11_63195978_rs3016865_ESE_RTN3_b1	14_91499257_rs4900092_tag_TRIP11_b1
11_63205701_rs11551944_no.sin_RTN3_b1	14_91505340_rs12587248_tag_TRIP11_b1
11_63211715_rs7943876_tag_RTN3_b1	14_91510819_rs1051340_no.sin_TRIP11_b1
11_63243962_rs542998_no.sin_RTN3_b1	14_91522261_rs17733414_tag_TRIP11_b1
11_63273401_rs17657473_tag_RTN3_b1	14_91527644_rs8017623_tag_TRIP11_b1
11_63283277_rs11551942_ESE_RTN3_b1	14_91529711_rs8007661_tag_TRIP11_b1
11_63793609_rs876064_tag_BAD_b1	14_91549050_rs11628699_tag_TRIP11_b1

14_91555634_rs7155279_tag_TRIP11_b1	19_2200477_rs10407022_no.sin_AMH_r1
14_91564183_rs17807271_tag_TRIP11_b1	19_10330975_rs12720356_no.sin_TYK2_r1
14_91567786_rs7157056_tag_TRIP11_b2	20_35404692_rs6017916_tag_SRC_b1
14_91569101_rs11621328_tag_TRIP11_b2	20_35410985_rs6017944_tag_SRC_b1
14_91574021_rs2235978_tag_TRIP11_b2	20_35417842_rs16986606_tag_SRC_b2
14_91575668_rs17127898_no.sin_TRIP11_r1	20_35432668_rs8126089_tag_SRC_b3
15_43200964_rs175088_TFBS_DUOX1_r1	20_35438751_rs6090576_tag_SRC_b4
15_43205544_rs16977750_TFBS_DUOX1_r2	20_35443071_rs6018199_tag_SRC_b4
15_43211181_rs17595239_tag_DUOX1_b1	20_35444271_rs911496_TFBS_SRC_b4
15_43211578_rs1706810_tag_DUOX1_b1	20_35455953_rs6018257_tag_SRC_b5
15_43213156_rs1648282_tag_DUOX1_b1	20_35462245_rs1570209_tag_SRC_b5
15_43222509_rs2292465_tag_DUOX1_b2	20_35466307_rs17785475_ESE_SRC_r1
15_43225829_rs1648305_tag_DUOX1_b2	21_33637569_rs2257167_no.sin_IFNAR1_r1
15_43231425_rs16939752_no.sin_DUOX1_r3	21_45120423_rs170962_tag_ITGB2_b1
15_43231810_rs1706804_ESE_DUOX1_b3	21_45126170_rs440555_tag_ITGB2_b1
15_43232791_rs1648304_tag_DUOX1_b3	21_45130589_rs684_tag_ITGB2_b1
15_43233448_rs2458236_no.sin_DUOX1_b3	21_45136241_rs235326_tag_ITGB2_b2
15_43244641_rs1648312_ESE_DUOX1_b3	21_45138385_rs235328_tag_ITGB2_b2
15_43244750_rs2292467_ESE_DUOX1_b3	21_45146246_rs1041457_tag_ITGB2_b3
15_43244811_rs1648311_ESE_DUOX1_b3	21_45146915_rs4607021_tag_ITGB2_b3
17_4551949_rs4790689_tag_ARRB2_b1	21_45147281_rs2026882_tag_ITGB2_b3
17_4555207_rs9895782_TFBS_ARRB2_r1	21_45152527_rs760462_tag_ITGB2_b4
17_23720820_rs2227741_no.sin_VTN_r1	21_45153345_rs760458_tag_ITGB2_b4
17_31215519_rs2291299_tag_CCL5_b1	21_45153843_rs760456_tag_ITGB2_b4
17_31221413_rs4795095_tag_CCL5_b1	21_45154168_rs2838734_tag_ITGB2_b4
17_31222706_rs1065341_ESE_CCL5_b1	21_45158230_rs3788150_tag_ITGB2_b4
17_35474634_rs1568400_tag_THRA_b1	21_45161993_rs3788151_tag_ITGB2_b5
17_37709372_rs9906989_tag_STAT3_b1	21_45168854_rs2838738_tag_ITGB2_b6
17_37719436_rs1053005_tag_STAT3_b1	22_20442794_rs3810608_tag_MAPK1_b1
17_37719618_rs1053004_ESE_STAT3_b1	22_20445004_rs6928_tag_MAPK1_b1
17_37719964_rs3744483_ESE_STAT3_b1	22_20445353_rs9340_tag_MAPK1_b1
17_37745206_rs2306580_tag_STAT3_b2	22_20520163_rs17759796_tag_MAPK1_b1
17_37748428_rs8069645_tag_STAT3_b2	22_20521585_rs8141815_tag_MAPK1_b1
17_37764060_rs9912773_tag_STAT3_b2	22_20525075_rs12172554_tag_MAPK1_b1
17_37779799_rs12949918_tag_STAT3_b2	22_20534793_rs8136867_TFBS_MAPK1_b1
17_37793938_rs17883843_ESE_STAT3_b2	22_20539139_rs9610470_tag_MAPK1_b1
17_37799793_rs16967738_TFBS_STAT3_b2	22_36394596_rs713835_tag_LGALS1_b1
17_37800443_rs17320971_TFBS_STAT3_b2	22_36396957_rs739139_TFBS_LGALS1_b1
17_45503724_rs2230390_no.sin_ITGA3_r1	22_36404380_rs9622682_tag_LGALS1_b1
17_45510424_rs2230392_no.sin_ITGA3_r2	22_39877954_rs20551_no.sin_EP300_r1
17_57384759_rs12941827_tag_THRAP1_b1	22_39904329_rs1046088_no.sin_EP300_r2
17_57402424_rs12451939_tag_THRAP1_b1	X_15269371_rs5935956_tag_FIGF_b1
17_57403256_rs6504077_tag_THRAP1_b1	X_15291030_rs6629030_tag_FIGF_b1
17_57490950_rs4422036_tag_THRAP1_b2	X_15291342_rs12011065_ESE_FIGF_b1
17_64014117_rs16972990_TFBS_PRKAR1A_r1	X_15312319_rs6632528_tag_FIGF_b2
17_64017451_rs4968898_tag_PRKAR1A_b1	X_15313272_rs5935959_tag_FIGF_b2
17_64019838_rs8080306_ESE_PRKAR1A_b1	X_15313948_rs5980152_tag_FIGF_b2
17_64029666_rs16973011_tag_PRKAR1A_b1	X_15316982_rs2317327_tag_FIGF_b2
17_64038281_rs1064757_ESE_PRKAR1A_b1	X_15317563_rs2027802_tag_FIGF_b2
17_64039429_rs9925_ESE_PRKAR1A_b1	X_15319641_rs4830943_TFBS_FIGF_b2
17_64045740_rs16973034_tag_PRKAR1A_b2	X_15323204_rs2071182_tag_FIGF_b2
18_44704974_rs7229639_tag_SMAD7_b1	X_15323550_rs2071177_tag_FIGF_b2
18_44707461_rs4939827_tag_SMAD7_b1	XY_1427370_rs17886756_no.sin_IL3RA_r1
18_44712225_rs9946510_tag_SMAD7_b2	XY_1457644_rs17883366_no.sin_IL3RA_r2
18_44712948_rs6507877_tag_SMAD7_b2	
18_44714901_rs2337106_tag_SMAD7_b3	
18_44715010_rs17186877_tag_SMAD7_b3	
18_44715784_rs7238442_tag_SMAD7_b3	
18_44717056_rs4939830_tag_SMAD7_b3	
18_44722249_rs1873191_tag_SMAD7_b4	
18_44722944_rs3764482_tag_SMAD7_b4	
18_44736068_rs2337143_TFBS_SMAD7_b5	
18_44736905_rs2337146_TFBS_SMAD7_b5	
18_46791487_rs2156010_tag_SMAD4_b1	
18_46821624_rs12968012_tag_SMAD4_b1	
18_46822269_rs10502913_tag_SMAD4_b1	

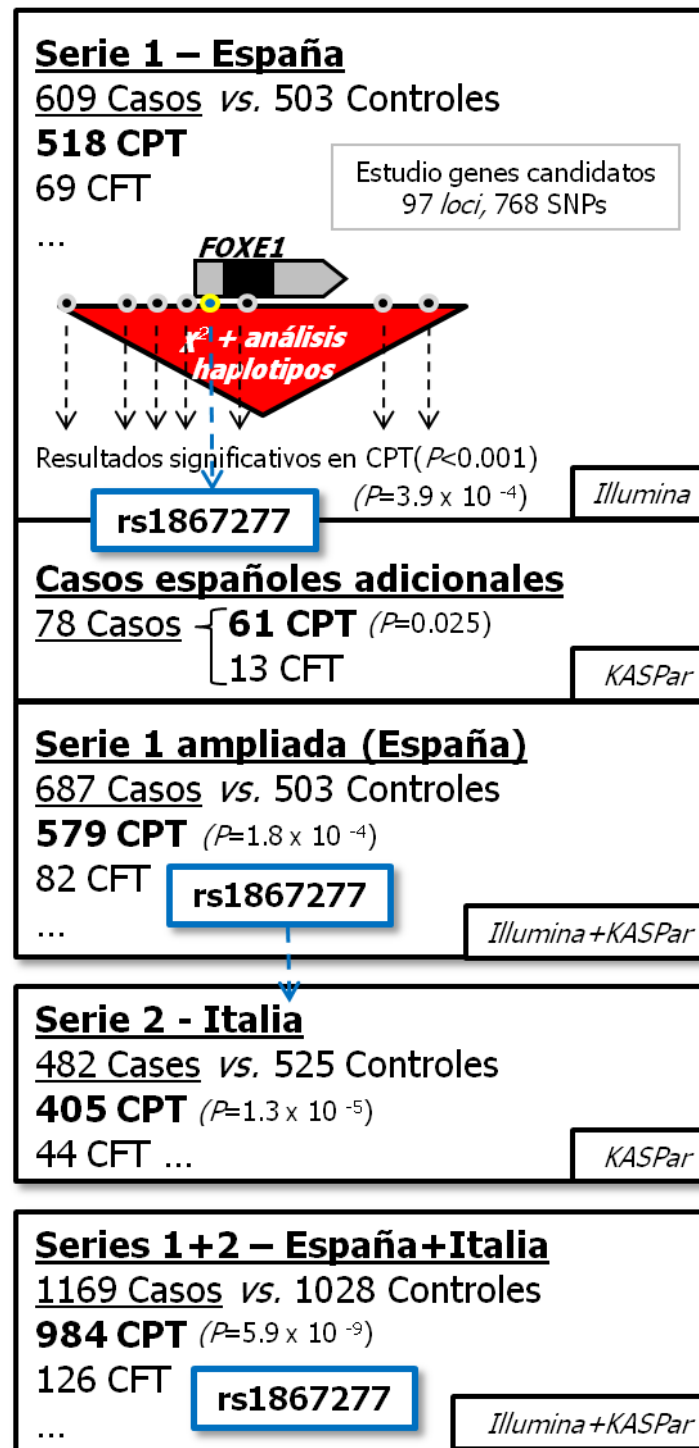
Los SNPs están ordenados de forma ascendente, primero por cromosoma y luego por posición dentro de cada cromosoma. Cada SNP aparece en el siguiente formato, separado por guiones bajos: cromosoma _ coordenada _ SNP rs _ función _ gen _ bloque LD o recombinación.

Tabla S2. Valores de desequilibrio de ligamiento (LD), expresados con el parámetro D' , para los SNPs del LD de *FOXE1* genotipados en nuestro trabajo mediante la plataforma Illumina, la variante funcional rs1867277 y el SNP rs965513 descrito por Gudmundsson y cols.

Dist (Kb) ^c	-59.8	-3.6	-1.8	-0.8	0	+1.1	+6.2	+8.7
SNP ^b (coordenada)	rs965513 (99,595,930)	rs894673 (99,652,091)	rs3758249 (99,653,961)	rs907577 (99,654,938)	rs1867277 (99,655,735)	rs3021526 (99,656,842)	rs874004 (99,661,939)	rs10119760 (99,664,423)
rs965513 (99,595,930)	CEU/ESP ^a	0.74 / NE	0.73 / NE	0.73 / NE	0.73 / NE	0.89 / NE	0.71 / NE	0.62 / NE
	rs894673 (99,652,091)		1.00 / 1.00	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00	1.00 / 0.98	0.96 / 0.99
		rs3758249 (99,653,961)		1.00 / 1.00	1.00 / 1.00	1.00 / 1.00	1.00 / 0.98	0.96 / 0.99
			rs907577 (99,654,938)		1.00 / 1.00	1.00 / 1.00	1.00 / 0.99	0.96 / 0.99
				rs1867277 (99,655,735)		1.00 / 1.00	1.00 / 0.99	0.96 / 0.99
					rs3021526 (99,656,842)		1.00 / 0.99	0.93 / 0.99
						rs874004 (99,661,939)		1.00 / 1.00
							rs10119760 (99,664,423)	

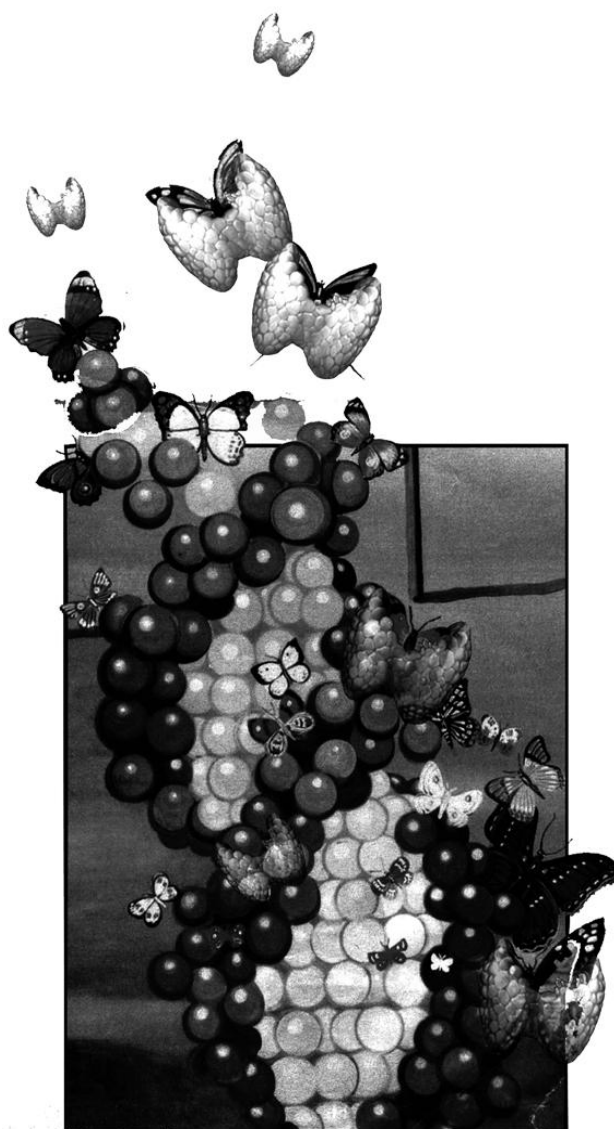
^a Los valores de D' se calcularon para la población HapMap-CEU (CEU) y para nuestros controles españoles incluidos en la serie 1 (ESP). Los valores de D' en población CEU para los SNPs rs1867277 y rs3021526 no estaban disponibles en las bases de datos y fueron inferidos a partir de sus SNPs correlacionados rs907577 y rs3021523, respectivamente. NE=no estudiado en esa población. ^b La variante funcional rs1867277 está resaltada en azul, y su tagSNP vecino, totalmente correlacionado con ella, se muestra en rojo. El SNP más significativo identificado por Gudmundsson y cols se muestra en verde. ^c Distancia física, medida en kilobases, respecto de la variante funcional rs1867277.

Figura S1. Diseño del estudio de asociación en dos pasos para *FOXE1*. El diagrama resume el número de muestras disponibles en el momento de la identificación de *FOXE1*, y complementa a la tabla R14 de la sección de Resultados. La serie 1 inicial, genotipada con el array de *Illumina*, se complementó con 78 casos adicionales, genotipados por KASPar. La serie 2 se corresponde con el grupo de casos y controles italianos descritos en el texto. En azul, se muestra la variante funcional rs1867277.



APÉNDICE II

Direcciones web



Catálogo de genes de baja penetrancia identificados en GWAS: <http://www.genome.gov/GWASudies/>

Colección de rutas metabólicas Biocarta: <http://www.biocarta.com/genes/index.asp>

Colección de rutas de la Enciclopedia de Genes y Genomas de Kyoto, *KEGG Pathways*: <http://www.genome.jp/kegg/pathway.html>

miRbase: <http://microrna.sanger.ac.uk/sequences/>

NCBI SNP database: <http://www.ncbi.nlm.nih.gov/snp>

Ensembl genome browser: <http://www.ensembl.org/index.html>

Proyecto Internacional HapMap: <http://hapmap.ncbi.nlm.nih.gov/>

Haploview: <http://www.broad.mit.edu/mpg/haploview>

PupaSuite: <http://pupasuite.bioinfo.cipf.es/>

F-SNP: <http://compbio.cs.queensu.ca/F-SNP/>

Tecnología de genotipado Illumina GoldenGate: http://www.illumina.com/technology/goldengate_genotyping_assay.ilmn

Tecnología de genotipado KASPar: <http://www.kbioscience.co.uk/download/KASP.swf>

Primer3: <http://biotools.umassmed.edu/bioapps/primer3> www.cgi

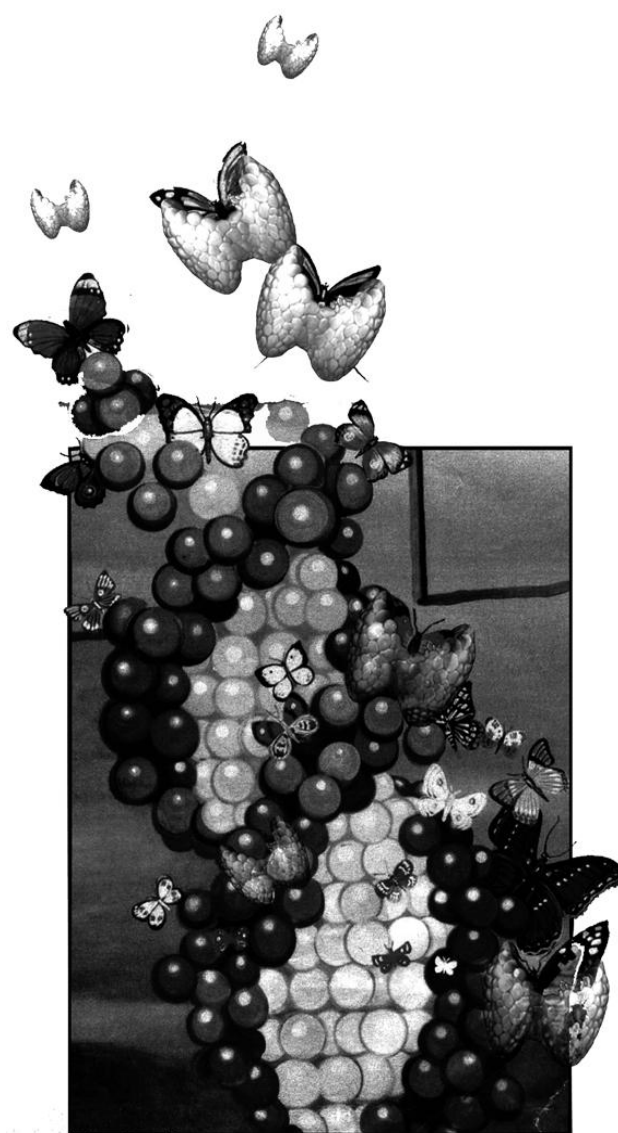
Proyecto R: <http://www.R-project.org>

Gene regulation: <http://www.gene-regulation.com>

APÉNDICE III

Publicaciones

derivadas de la tesis



A continuación, se adjunta una copia de las siguientes publicaciones, derivadas de esta tesis:

1: Landa I, Montero-Conde C, Malanga D, De Gisi S, Pita G, Leandro-García LJ, Inglada-Pérez L, Letón R, De Marco C, Rodríguez-Antona C, Viglietto G, Robledo M.

Allelic variant at -79 (C>T) in *CDKN1B* (p27^{Kip1}) confers an increased risk of thyroid cancer and alters mRNA levels.

Endocr Relat Cancer. 2010 Jun 1;17(2):317-28. Print 2010 Jun. PubMed PMID: 20075119.

2: Landa I, Ruiz-Llorente S, Montero-Conde C, Inglada-Pérez L, Schiavi F, Leskelä S, Pita G, Milne R, Maravall J, Ramos I, Andía V, Rodríguez-Poyo P, Jara-Albarrán A, Meoro A, del Peso C, Arribas L, Iglesias P, Caballero J, Serrano J, Picó A, Pomares F, Giménez G, López-Mondéjar P, Castello R, Merante-Boschin I, Pelizzo MR, Mauricio D, Opocher G, Rodríguez-Antona C, González-Neira A, Matías-Guiu X, Santisteban P, Robledo M.

The variant rs1867277 in *FOXE1* gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors.

PLoS Genet. 2009 Sep;5(9):e1000637. Epub 2009 Sep 4. PubMed PMID: 19730683; PubMed Central PMCID: PMC2727793.

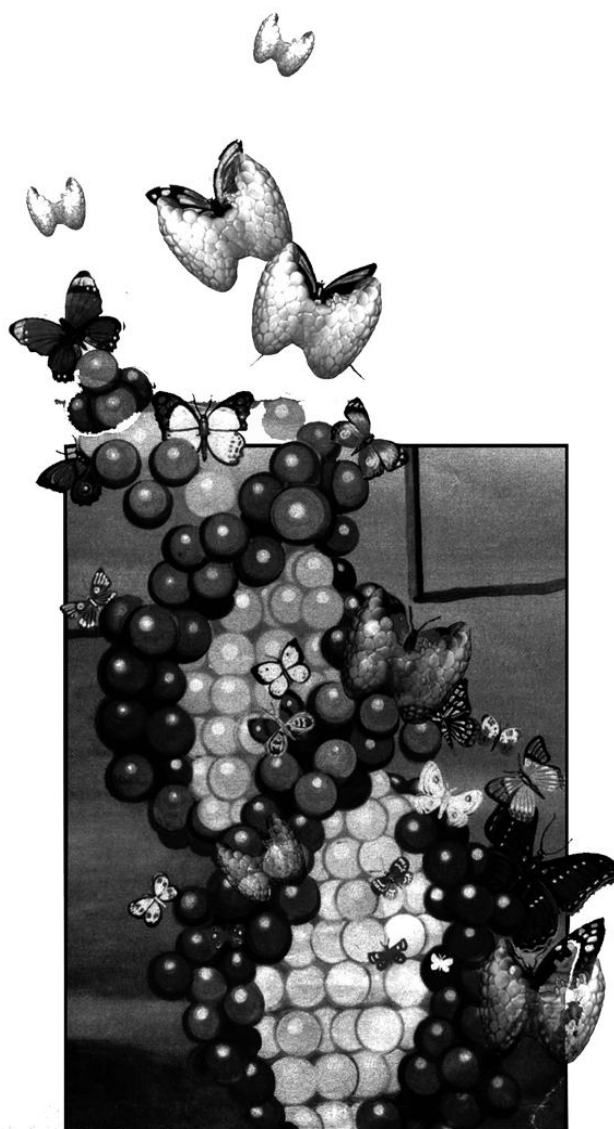
3: Landa I & Robledo M.

Association studies in thyroid cancer susceptibility. Are we on the right track?

J Mol Endocrinol. 2011 [in press].

APÉNDICE IV

Otras publicaciones



1. Mol Endocrinol. 2010 Dec;24(12):2382-91. Epub 2010 Oct 27.

Research resource: Transcriptional profiling reveals different pseudohypoxic signatures in SDHB and VHL-related pheochromocytomas.

López-Jiménez E, Gómez-López G, Leandro-García LJ, Muñoz I, Schiavi F, Montero-Conde C, de Cubas AA, Ramires R, Landa I, Leskelä S, Maliszewska A, Inglada-Pérez L, de la Vega L, Rodríguez-Antona C, Letón R, Bernal C, de Campos JM, Díez-Tascón C, Fraga MF, Boullosa C, Pisano DG, Opocher G, Robledo M, Cascón A.

The six major genes involved in hereditary susceptibility for pheochromocytoma (PCC)/paraganglioma (PGL) (RET, VHL, NF1, SDHB, SDHC, and SDHD) have been recently integrated into the same neuronal apoptotic pathway where mutations in any of these genes lead to cell death. In this model, prolyl hydroxylase 3 (Egln3) abrogation plays a pivotal role, but the molecular mechanisms underlying its inactivation are currently unknown. The aim of the study was to decipher specific alterations associated with the different genetic classes of PCCs/PGLs. With this purpose, 84 genetically characterized tumors were analyzed by means of transcriptional profiling. The analysis revealed a hypoxia-inducible factor (HIF)-related signature common to succinate dehydrogenase (SDH) and von Hippel-Lindau (VHL) tumors, that differentiated them from RET and neurofibromatosis type 1 cases. Both canonical HIF-1 α and HIF-2 α target genes were overexpressed in the SDH/VHL cluster, suggesting that a global HIF deregulation accounts for this common profile. Nevertheless, when we compared VHL tumors with SDHB cases, which often exhibit a malignant behavior, we found that HIF-1 α target genes showed a predominant activation in the VHL PCCs. Expression data from 67 HIF target genes was sufficient to cluster SDHB and VHL tumors into two different groups, demonstrating different pseudo-hypoxic signatures. In addition, VHL-mutated tumors showed an unexpected overexpression of Egln3 mRNA that did not lead to significantly different Egln3 protein levels. These findings pave the way for more specific therapeutic approaches for malignant PCCs/PGLs management based on the patient's genetic alteration.

2. Pharmacogenomics J. 2010 Mar 9. [Epub ahead of print]

Polymorphisms in cytochromes P450 2C8 and 3A5 are associated with paclitaxel neurotoxicity.

Leskelä S, Jara C, Leandro-García LJ, Martínez A, García-Donas J, Hernando S, Hurtado A, Vicario JC, Montero-Conde C, Landa I, López-Jiménez E, Cascón A, Milne RL, Robledo M, Rodríguez-Antona C.

Neurotoxicity is one of the most relevant dose-limiting toxicities of the anticancer drug paclitaxel. It exhibits substantial interindividual variability of unknown molecular basis, and represents one of the major challenges for the improvement of paclitaxel therapy. The extensive variability in paclitaxel clearance and metabolism lead us to investigate the association between polymorphisms in paclitaxel elimination pathway and neurotoxicity. We selected 13 relevant polymorphisms in genes encoding paclitaxel metabolizing enzymes (CYP2C8, CYP3A4 and CYP3A5) and transporters (organic anion transporting polypeptide (OATP) 1B1, OATP1B3 and P-glycoprotein) and genotyped them in 118 Spanish cancer patients treated with paclitaxel. After adjusting for age and treatment schedule, CYP2C8 Haplotype C and CYP3A5*3 were associated with protection (hazard ratio (HR) (per allele)=0.55; 95% confidence interval (CI)=0.34-0.89; P=0.014 and HR (per allele)=0.51; 95%CI=0.30-0.86; and P=0.012, respectively) and CYP2C8*3 with increased risk (HR (per allele)=1.72; 95%CI=1.05-2.82; and P=0.032). In each case, the allele causing increased paclitaxel metabolism was associated with increased neurotoxicity, suggesting an important role for metabolism and hydroxylated paclitaxel metabolites. We estimated the HR per paclitaxel-metabolism increasing allele carried across the three polymorphisms to be HR=1.64 (95% CI=1.26-2.14; P=0.0003). The results for P-glycoprotein were inconclusive, and no associations were observed for the other genes studied. The incorporation of this genetic data in treatment selection could help to reduce neurotoxicity events, thereby individualizing paclitaxel pharmacotherapy. These results warrant validation in independent series.

3. Cytoskeleton (Hoboken). 2010 Apr;67(4):214-23.

Tumoral and tissue-specific expression of the major human beta-tubulin isotypes.

Leandro-García LJ, Leskelä S, Landa I, Montero-Conde C, López-Jiménez E, Letón R, Cascón A, Robledo M, Rodríguez-Antona C.

The beta-tubulins are microtubule components encoded by a multigene family, which produces slightly different proteins with complex expression patterns. Several widely used anticancer drugs base their activity on beta-tubulin binding, microtubule dynamics alteration, and cell division blockage. The expression of these drug targets in tumoral and normal cells could be of crucial importance for therapy outcome, unfortunately, the complex beta-tubulin expression patterns have been poorly characterized in human. In this study, we developed a quantitative RT-PCR technique that accurately determines the mRNA expression of the eight human beta-tubulin isotypes, encoding class I, IIa, IIb, III, IVa, IVb, V, and VI and applied it to 21 nontumoral tissues and 79 tumor samples belonging to seven cancer types. In the nontumoral tissues, we found that, overall, TUBB (I), TUBB2C (IVb), and TUBB6 (V) were ubiquitous, TUBB1(VI) was hematopoietic cell-specific, and TUBB2A (IIa), TUBB2B (IIb), TUBB3 (III), and TUBB4 (IVa) had high expression in brain; however, the contribution of the different isotypes to the total beta-tubulin content varied for each tissue and had a complex pattern. In tumoral tissues, most isotypes exhibited an altered expression in specific tumor types or related to tumoral characteristics. In general, TUBB3 showed a great increase in expression while TUBB6 expression was largely decreased in most tumors. Thus, normal tissues showed a complex beta-tubulin isotype distribution, which could contribute to the toxicity profile of the microtubule-binding drugs. In addition, the specific isotypes significantly altered in tumors might represent markers for drug response.

4. Endocr Relat Cancer. 2010 Jan 29;17(1):7-16. Print 2010 Mar.

Overexpression and activation of EGFR and VEGFR2 in medullary thyroid carcinomas is related to metastasis.

Rodríguez-Antona C, Pallares J, Montero-Conde C, Inglada-Pérez L, Castelblanco E, Landa I, Leskelä S, Leandro-García LJ, López-Jiménez E, Letón R, Cascón A, Lerma E, Martín MC, Carralero MC, Mauricio D, Cigudosa JC, Matias-Guiu X, Robledo M.

Therapeutic options for patients with metastatic medullary thyroid carcinoma (MTC) are limited due to lack of effective treatments. Thus, there is a need to thoroughly characterize the pathways of molecular pathogenesis and to identify potential targets for therapy in MTC. Since epidermal growth factor receptor (EGFR) seems to play a crucial role for RET activation, a key feature of MTCs, and several promising EGFR/vascular endothelial growth factor receptor 2 (VEGFR2)-targeted drugs have been developed, the present study was designed to investigate whether these proteins are altered in MTCs. We used a well-characterized series of 153 MTCs to evaluate EGFR activation by sequencing and FISH analysis, and to perform EGFR and VEGFR2 immunohistochemistry. EGFR tyrosine kinase domain mutations were not a feature of MTCs; however, EGFR polysomy and a strong EGFR expression were detected in 15 and 13% of the tumors respectively. Interestingly, EGFR was significantly overexpressed in metastases compared with primary tumors (35 vs 9%, $P=0.002$). We also studied whether specific RET mutations were associated with EGFR status, and found a decrease in EGFR polysomies ($P=0.006$) and a tendency towards lower EGFR expression for the most aggressive RET mutations (918, 883). Concerning VEGFR2, metastasis showed a higher expression than primary tumors ($P=2.8 \times 10^{-8}$). In this first study investigating the relationship between EGFR, RET, and VEGFR2 in a large MTC series, we found an activation of EGFR and VEGFR2 in metastasis, using both independent and matched primary/metastasis samples. This suggests that some MTC patients may benefit from existing anti-EGFR/VEGFR2 therapies, although additional preclinical and clinical evidence is needed.

5. Horm Metab Res. 2009 Sep;41(9):672-5. Epub 2009 Apr 2.

Rationalization of genetic testing in patients with apparently sporadic pheochromocytoma/paraganglioma.

Cascón A, López-Jiménez E, Landa I, Leskelä S, Leandro-García LJ, Maliszewska A, Letón R, de la Vega L, García-Barcina MJ, Sanabria C, Alvarez-Escolá C, Rodríguez-Antona C, Robledo M.

Hereditary susceptibility to pheochromocytoma (PCC) and paraganglioma (PGL) represents a very complex genetic scenario. It has been reported that the absence of familial antecedents of the disease does not preclude the existence of a mutation affecting any of the five major susceptibility genes. In fact, 11-24% of apparently sporadic cases (without familial or syndromic antecedents) harbor an unexpected germline mutation, but we do not know what is happening in "truly apparently" sporadic patients (i.e., apparently sporadic cases diagnosed with only one tumor). In the present study, we have analyzed 135 apparently sporadic patients developing a single tumor for the five major susceptibility genes: VHL, RET, SDHB, SDHC, and SDHD. Fourteen percent of cases were found to harbor a germline mutation, and only 2.2% of patients were older than 45 years at onset. By taking into account the tumor location and a threshold age at onset of 45 years, we propose a rational scheme for genetic testing. Analyzing VHL and RET genes would be recommended only in young patients developing a single PCC. On the other hand, genetic testing of SDHD should be done in all patients developing an extra-adrenal tumor before the age of 45, and SDHC could be the responsible gene in cases developing a single head and neck tumor, independently of age. Finally, the analysis of SDHB should always be performed because of its association to malignancy and the low penetrance of mutations affecting this gene.

6. Anal Biochem. 2009 Jun 1;389(1):74-6. Epub 2009 Mar 20.

Determination of CYP2D6 gene copy number by multiplex polymerase chain reaction analysis.

Leandro-García LJ, Leskelä S, Montero-Conde C, Landa I, López-Jimenez E, Letón R, Seeringer A, Kirchheiner J, Cascón A, Robledo M, Rodríguez-Antona C.

Cytochrome P450 2D6 (CYP2D6) copy number variation (CNV) influences the metabolism of 15-25% of clinical drugs. Here we describe a novel multiplex polymerase chain reaction (PCR) analysis method that accurately detects CYP2D6 CNV and CYP2D6*9 allele. It includes the amplification of 2 CYP2D6 and 7 control (AQP1, CYP3A4, MDR1, and SDHB) fluorescent PCR products that are separated on a capillary sequencer and normalized using reference samples. The technique was validated using 27 PCR-restriction fragment length polymorphism (RFLP) pregenotyped samples and further tested in 75 Caucasian samples. The method assigns the correct CYP2D6 copy number, independent of already characterized CYP2D6 single nucleotide polymorphisms (SNPs), and could easily be applied to clinical samples.

7. J Clin Endocrinol Metab. 2009 May;94(5):1701-5. Epub 2009 Mar 3.

Genetics of pheochromocytoma and paraganglioma in Spanish patients.

Cascón A, Pita G, Burnichon N, Landa I, López-Jiménez E, Montero-Conde C, Leskelä S, Leandro-García LJ, Letón R, Rodríguez-Antona C, Díaz JA, López-Vidriero E, González-Neira A, Velasco A, Matias-Guiu X, Gimenez-Roqueplo AP, Robledo M.

CONTEXT: The presence of familial history in pheochromocytoma/paraganglioma patients, including syndromic antecedents, leads in the majority of cases to a positive genetic testing for mutations in one of the major susceptibility genes described so far. Furthermore, it has been reported that in the absence of familial antecedents, about 11-24% of patients also carry a mutation in one of these related genes. In these cases, other clinical aspects like bilaterality, multiplicity, location of the tumors, or age at onset can help to recognize the underlying genes involved. OBJECTIVE: The objective of the study was to discuss clinical criteria helpful in the genetic diagnosis, placing special emphasis on apparently sporadic cases. DESIGN: Two hundred thirty-seven nonrelated probands were analyzed for the major susceptibility genes: VHL, RET, SDHB, SDHC, and SDHD. Genetic characterization included both point mutation analysis and gross deletions in the SDH genes performed by multiplex PCR. RESULTS: As expected, all syndromic probands were genetically diagnosed with a mutation affecting either RET or VHL. A total of 79.1% (19 of 24) and 18.4% (31 of 168) of patients presenting with either nonsyndromic familial antecedents or apparently sporadic presentation were found to carry a mutation in one of the susceptibility genes. Finally, we found a Spanish founder effect for two mutations: SDHB c.166_170delCCTCA and SDHD c.129G>A. CONCLUSIONS: Germline mutations are rare in apparently sporadic probands diagnosed after age 40 yr (3.9% in our series) and mainly involve SDHB. Therefore, we recommend prioritizing SDHB genetic testing in patients developing isolated tumors at any age, especially those with extraadrenal location or malignant behavior.

8. Clin Endocrinol (Oxf). 2008 Dec;69(6):906-10. Epub 2008 Aug 4.

SDHC mutation in an elderly patient without familial antecedents.

López-Jiménez E, de Campos JM, Kusak EM, Landa I, Leskelä S, Montero-Conde C, Leandro-García LJ, Vallejo LA, Madrigal B, Rodríguez-Antona C, Robledo M, Cascón A.

Head and neck paragangliomas are usually asymptomatic and benign tumours arising mainly from the carotid body and the vagal, tympanic or jugular glomus. The majority of patients develop sporadic masses, and around 30% of cases harbor germline mutations in one of the succinate dehydrogenase genes: SDHB, SDHC or SDHD. In these hereditary cases, the presence of familial antecedents of the disease, multiplicity/bilaterality, young age at onset, and more recently, presence of gastrointestinal stromal tumours, are main factors to be considered. Here we describe a new mutation (c.256-257insTTT) affecting the SDHC gene in a 60-year-old-patient with a single head and neck paraganglioma, and without familial antecedents of the disease. In silico splice site analysis showed that this variant created a cryptic splice acceptor site and loss of heterozygosity (LOH) supported the pathogenic role of the mutation. Control population analyses did not detect this variant but revealed a novel SDHC polymorphism that exhibited a frequency of 0.3% (3/1020). This latter finding highlights the importance of assessing the clinical relevance of variants of unknown significance by means of analysing sufficient controls. Despite having found a germline mutation in an older, apparently sporadic patient, we consider that the high costs of analyzing all susceptibility genes related to the disease support the recommendation of screening for mutations only in patients fulfilling the above criteria.

9. J Med Genet. 2008 Apr;45(4):233-8. Epub 2007 Dec 5.

Molecular characterisation of a common SDHB deletion in paraganglioma patients.

Cascón A, Landa I, López-Jiménez E, Díez-Hernández A, Buchta M, Montero-Conde C, Leskelä S, Leandro-García LJ, Letón R, Rodríguez-Antona C, Eng C, Neumann HP, Robledo M.

BACKGROUND: Hereditary susceptibility to familial paraganglioma syndromes is mainly due to mutations in one of six genes, including three of the four genes encoding the subunits of the mitochondrial succinate dehydrogenase complex II. Although prevalence, penetrance and clinical characteristics of patients carrying point mutations affecting the genes encoding succinate dehydrogenase have been well studied, little is known regarding these clinical features in patients with gross deletions. Recently, we found two unrelated Spanish families carrying the previously reported SDHB exon 1 deletion, and suggested that this chromosomal region could be a hotspot deletion area. **METHODS:** We present the molecular characterisation of this apparently prevalent mutation in three new families, and discuss whether this recurrent mutation is due either to the presence of a founder effect or to a hotspot. **RESULTS:** The breakpoint analysis showed that all Iberian Peninsular families described harbour the same exon 1 deletion, and that a different breakpoint junction segregates in an affected French pedigree. **CONCLUSIONS:** After haplotyping the SDHB region, we concluded that the deletion detected in Iberian Peninsular people is probably due to a founder effect. Regarding the clinical characteristics of patients with this alteration, it seems that the presence of gross deletions rather than point mutations is more likely related to abdominal presentations and younger age at onset. Moreover, we found for the first time a patient with neuroblastoma and a germline SDHB deletion, but it seems that this paediatric neoplasia in a pheochromocytoma family is not a key component of this disease.

10. Endocr Relat Cancer. 2007 Sep;14(3):645-54.

Cytochrome P450 3A5 is highly expressed in normal prostate cells but absent in prostate cancer.

Leskelä S, Honrado E, Montero-Conde C, Landa I, Cascón A, Letón R, Talavera P, Cózar JM, Concha A, Robledo M, Rodríguez-Antona C.

Testosterone is essential for the growth and function of the luminal prostate cells, but it is also critical for the development of prostate cancer, which in the majority of the cases derives from luminal cells. Cytochrome P450 3A (CYP3A) enzymes hydroxylate testosterone and dehydroepiandrosterone to less active metabolites, which might be the basis for the association between CYP3A polymorphisms and prostate cancer. However, it is unknown whether the CYP3A enzymes are expressed at relevant levels in the prostate and which polymorphisms could affect this tissue-specific CYP3A activity. Thus, we measured CYP3A4, CYP3A5, CYP3A7, and CYP3A43 mRNA in 14 benign prostatic hyperplasias and ten matched non-tumoral/tumoral prostate samples. We found that CYP3A5 mRNA in non-tumoral prostate tissue was 10% of the average amount of liver samples, whereas the expression of the other CYP3A genes was much lower. Similarly to liver, CYP3A5*3 polymorphism decreased CYP3A5 mRNA content 13-fold. CYP3A5 protein was detected in non-tumoral prostate microsomes by western blot, and immunohistochemistry (IHC) localized CYP3A5 exclusively in the basolateral prostate cells. In contrast to the normal tissue, IHC and RT-PCR showed that tumoral tissue lacked CYP3A5 expression. In conclusion, prostate basolateral cells express high levels of CYP3A5 which dramatically decrease in tumoral tissue. This finding supports an endogenous function of CYP3A5 related to the metabolism of intra-prostatic androgens and cell growth, and that polymorphisms affecting CYP3A5 activity may result in altered prostate cancer risk and aggressiveness.

11. Cancer Res. 2007 Oct 1;67(19):9561-7.

Association study of 69 genes in the ret pathway identifies low-penetrance loci in sporadic medullary thyroid carcinoma.

Ruiz-Llorente S, Montero-Conde C, Milne RL, Moya CM, Cebrián A, Letón R, Cascón A, Mercadillo F, Landa I, Borrego S, Pérez de Nanclares G, Alvarez-Escolá C, Díaz-Pérez JA, Carracedo A, Urioste M, González-Neira A, Benítez J, Santisteban P, Dopazo J, Ponder BA, Robledo M; Medullary Thyroid Carcinoma Clinical Group.

To date, few association studies have been done to better understand the genetic basis for the development of sporadic medullary thyroid carcinoma (sMTC). To identify additional low-penetrance genes, we have done a two-stage case-control study in two European populations using high-throughput genotyping. We selected 417 single nucleotide polymorphisms (SNP) belonging to 69 genes either related to RET signaling pathway/functions or involved in key processes for cancer development. TagSNPs and functional variants were included where possible. These SNPs were initially studied in the largest known series of sMTC cases (n = 266) and controls (n = 422), all of Spanish origin. In stage II, an independent British series of 155 sMTC patients and 531 controls was included to validate the previous results. Associations were assessed by an exhaustive analysis of individual SNPs but also considering gene- and linkage disequilibrium-based haplotypes. This strategy allowed us to identify seven low-penetrance genes, six of them (STAT1, AURKA, BCL2, CDKN2B, CDK6, and COMT) consistently associated with sMTC risk in the two case-control series and a seventh (HRAS) with individual SNPs and haplotypes associated with sMTC in the Spanish data set. The potential role of CDKN2B was confirmed by a functional assay showing a role of a SNP (rs7044859) in the promoter region in altering the binding of the transcription factor HNF1. These results highlight the utility of association studies using homogeneous series of cases for better understanding complex diseases.

